Real-time Prediction of Hospital Length of Stay for Surgical Patients: a Data-drive Approach

A thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering

Xincheng Tan

M.E. Degree Candidate in Computational Science and Engineering

Faculty Advisor: Mauricio Santillana

Harvard University School of Engineering and Applied Science

Cambridge MA

May 13, 2022

Contents

List of Figures

List of Tables

Abstract

The ability to predict the length of stay (LOS) for patients who will undergo a surgical procedure, at the moment the surgery is scheduled, upon admission, and once the procedure has taken place, is a critical task to improve resource allocation at hospitals. We present a study that leverages patient information from electronic health records (EHRs), available at the time of prediction, to build various supervised learning models to predict the LOS for inpatient visits across all hospital service lines. We explore a wide range of machine learning classifiers on a feature set combining patient demographics, health status and the complexity of surgical procedure features. Our models outperform surgeon's predictions by over 3% in the overall accuracy with comparable mean absolute error. Ultimately, we propose a decision support system that integrates both expert judgements and machine intelligence tailored for multiple surgical specialties.

Keywords hospital length of stay, bed management, supervised learning, health informatics, decision support

Acknowledgements

This research is supported by the Institute of Applied Computational Science Fellowship at Harvard University. We gratefully acknowledge the computing resources provided by IACS at Harvard University, and the dataset and communication resources by Boston Children's Hospital. We also thank Jay Berry, Peter Hong, Lynne Ferrari, Alyssa Teves, Gabor Asztalos, Anna Kordun, Steven Staffa, Anna Desmarais, Rachel Difazio, Izabela Leahy, Joshua Merrill for project scoping, in-depth discussion and timely feedback, and Prof. Mauricio Santillana, Andre Nguyen and Jack Wang for technical mentorship and support.

Chapter 1

Introduction

1.1 Background

The primary goal of healthcare providers is to deliver high-quality and timely patient care while minimizing hospital cost. It requires precise clinical diagnosis and effective treatments from the clinical teams, as well as efficient management of hospital resources from the administrative department. To optimize resource utilization, the hospital admin staff needs to efficiently allocate medical facilities, nursing care and bed units for inpatients. In particular, hospital bed capacity is crucial but limited resource, so an optimal bed management strategy aims at shorter patient wait time and high bed occupancy rate [1]. There are three major categories of inpatients who demand using hospital beds overnight: surgical, medical and behavioral health patients. Surgical patients undergo planned or emergency surgical procedures; medical patients are diagnosed with disease or illness which require medical intervention; behavioral health patients require hospital-level psychiatric aides.

At Boston Children's Hospital, there are over 20,000 inpatient visits every year, among which over 5,000 are surgical patients who undergo scheduled surgeries and anticipate overnight hospital stay. The ability to properly manage beds for the scheduled procedures is imperative to maintaining a healthy throughput of patients and delivering better patient care.

Hospital length of stay (LOS) provides critical information to developing a professional bed scheduling system. Without any prior knowledge of a patient's expected LOS at the time of scheduling, schedulers may book longer bed days than needed, which incurs resource waste and hampers scheduling upcoming inpatient visits, or have to extend the bed occupancy unexpectedly, which would delay the care delivery to incoming patients. In addition to the expected LOS at scheduling time, the real-time LOS is also helpful to capacity management. Researchers have found that the elapsed length of stay is a strong predictor of early discharge (by 2 p.m.) versus midnight discharge, which is a key decision to help clinicians organize and prioritize their daily tasks [2].

Predictive modeling of LOS has gained increasing momentum with the digitization of health records. Structured electronic health records enable operational researchers to derive insights from a large quantity of patient data not only from a clinical angle, but also from an operational viewpoint. Fig1.1 shows the flow of the critical hospital events associated with each surgical inpatient visit across time in Boston Children's Hospital. At the time of surgical request, patient's demographics, medication history, diagnoses and primary procedure information is available. However, depending on the type of primary procedure, the patient might undergo auxiliary procedures involving surgical teams from other specialties. A full list of procedures (in CPT coding) is only available 24 hours after surgical request. In this study, we focus on predicting LOS prior to patient admission and do not concern with dynamic predictions throughout the full patient visit.

1.2 Objectives

In this study, we investigate a variety of machine learning models to infer LOS for patients who undergo planned surgeries across all hospital service lines. We develop and deploy a machine learning framework that predicts LOS before admission–within 24 hours after surgical request, at the time of admission and after the surgical procedure completes. We formulate this problem as a classification task and applied various model families with detailed analysis on their performance compared to surgeon's prediction. To better assist surgeons in operational decision making, we further quantify the prediction uncertainty leveraging surgeon's prediction and the surgical cohort categorization.

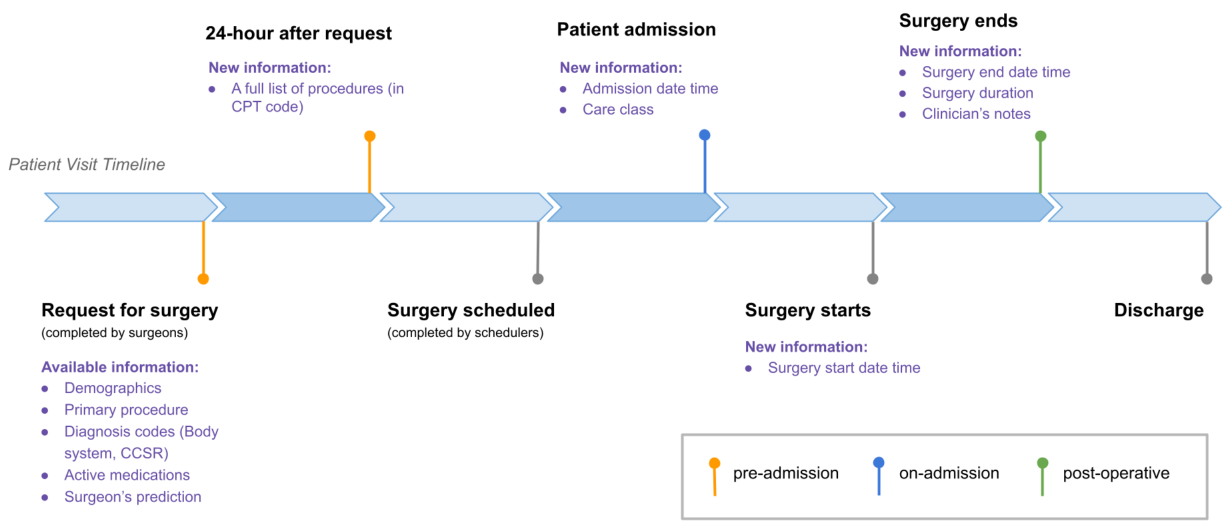


Figure 1.1: Diagram of a typical inpatient hospital visit events and the associated information available up to each time point.

1.3 Related Works

A plethora of works aim to predict patient LOS under various contexts. Existing LOS prediction studies mainly differ in three aspects: patient population, predictive task formulation and timing of prediction. A diverse set of modeling approaches have been investigated among the combinations of all three aspects. Their performance has a high variability across studies and is largely dependent on the specifics in task formulation and the actual dataset.

1.3.1 Patient Population

The medical literature has defined a number of axes to categorize the patient population. Medical specialities, diagnosis types and admission units (emergency department, ICU or regular wards) are the most common groupers in stratifying the patient population. Different categories of patients exhibit heterogeneous statistical properties such as skewness and kurtosis in their LOS distribution [3]. In a 18-year retrospective study of hospitalizations in a tertiary healthcare center in Mexico, Marfil-Garza *et al.* found that emergency and surgical admissions were associated with an increased risk of prolonged length of stay (LOS $\geq$ 34 days), and patients with certain diagnosis, such as bone marrow transplant, fungal and bacterial infections, hematological neoplasms, etc., had an increased risk for prolonged LOS [4].

To account for such variation in patient population, several LOS prediction studies included in their predictive modeling only the patients who were diagnosed with specific conditions or underwent procedures within specific medical specialties. In predicting LOS for three types of cardiology patients, Tsai *et al.* compared linear regression and a 3-layer artificial neural network (ANN), and find that linear regression yields higher accuracy ratings [5]. Gao *et al.* incorporated disease-specific information into random forest models to infer LOS for obstetrics patients, obtaining a prediction accuracy of 49.3% [6]. To predict LOS of congestive heart failure patients at admission time, Turgeman *et al.* compared generalized linear model, support vector machine, CART tree, Chi-squared automatic interaction detection tree, neural network and Cubist models, achieving a best mean absolute error of 1 day [7].

Instead of modeling on a particular patient cohort, a few LOS studies incorporated the clinical variables directly in their modeling over a general patient population. Bert *et al.* treated diagnoses variables as binary indicators in the regression analysis, and discovered that patients under ED admission or diagnosed with neoplasia had significantly longer LOS [8]. Other studies leverage clinical expertise to encode clinical information as risk score variables in their modeling approach. In predicting the prolonged LOS for general surgical patients, Chuang *et al.* leveraged surgeon-assigned preoperative risk assessment scores and the number of comorbidities per patient visit in the random forest model, achieving a best performance of 0.92 in area under the curve [9].

Previous studies adopted dichotomous representations of clinical type variables, but none of them investigated their impact on the model performance over the same dataset. In this study, we compare training a *global* machine learning model on the full patient population with training a set of *local* models, one for each patient cohort grouped by clinical categories.

1.3.2 Task Formulation

Length of stay is a continuous quantity measured in time units, but it can be treated as a discrete variable depending on the predictive task. The predicted LOS outcome can mainly be formulated as:

1. a continuous variable in the unit of hour or day
2. a binary variable according to a human-defined threshold (i.e. whether LOS > threshold or not)
3. a categorical variable that represents a time range

When LOS is modeled as a continuous variable, regression methods are applied, oftentimes with regularization or in the context of tree-based and ensemble models. Performance is mostly assessed with R^2, the mean squared error (MSE) or the mean absolute error (MAE) [10]. In predicting length of stay for ICU patients in the Medical Information Mart for Intensive Care (MIMIC) dataset, Sanjay *et al.* benchmarked a super-learner model and three deep neural networks (feedforward neural network, recurrent neural network and multimodal deep learning model), and reported that multimodal deep learning method achieved the lowest MSE of 36,338 in hours [11]. In another study to predict LOS for emergency and accident admissions, Stone *et al.* compared 10 machine learning techniques such as fuzzy nearest neighbor, Naive Bayes and variants of fuzzy-rough learners, and discovered that Zero Rule (majority class) method yielded the lowest RMSE of 11 days [12].

When defined as a discrete variable, the LOS outcome is more often predicted via classification models, such as support vector machines, tree-based ensemble models, neural networks, etc. In the discrete outcome setting, the most common evaluation metrics include accuracy, AUROC (area under the receiver operating characteristic curve) and sensitivity [10].

Studies that consider LOS as a binary outcome typically aim to distinguish between short versus long length of stay. In predicting prolonged LOS (> 4 days) in a pediatric ICU unit, Castiñeira *et al.* combined continuous vital sign information with static patient health records and applied hierarchical classifiers based on gradient boosting models, achieving the best performance of 0.9 AUROC [13]. In another study of clinical machine learning tasks on the MIMIC dataset, Nestor *et al.* leveraged logistic regression, random forest and recurrent neural networks in predicting whether an admitted ICU patient would stay in ICU for more than 3 days, obtaining a best average AUROC of 0.85 via random forest [14].

To increase the outcome resolution, LOS can be further treated as a categorical variable of time ranges. In this context, researchers leverage multi-class classification models to predict the time range of patient's LOS. The definition of such time ranges is arbitrary and at the sole discretion of researchers and clinical experts. In predicting the LOS for ICU patients, Abd-Elrazek *et al.* divided LOS into 3 classes: ≤ 2 days, 3–7 days, and > 7 days. They applied the re-sampling technique, bootstrap aggregating (bagging), based on six machine learning techniques (k-nearest neighbors, fuzzy rules, classification tree, regression tree, random forest and tree bagger) and obtained the best performance in classification tree with an average accuracy of 90% and an average recall and precision of 85% [15]. In another study of modeling LOS for cardiac patients, Daghistan *et al.* divided the LOS outcome into three categories: short (< 3 days), intermediate (3–5 days) and long (> 5 days), and reported that random forest model achieved a best performance of 0.94 in AUROC [16].

Despite the exhaustive modeling efforts, none of the previous works carried out systematic diagnostics on the model prediction, nor did they quantify the uncertainty of the prediction. In this paper, we present a pilot scheme based on clinical information to help domain experts as well as data scientists better understand the model's strength and weakness.

1.3.3 Prediction Timing

Timing of LOS prediction is another imperative aspect related to its modeling. At each stage along the patient visit timeline, decision makers, models or humans, are somewhat constrained to the patient information available at that time. For example, although several studies have found that lab test results and vital sign signals are predictive of LOS [13, 9, 17], these clinical information are only available *after* admission. Thus, the prediction models at the pre-admission stage cannot utilize these features in their decision making. In a systematic review of 74 LOS prediction studies from 1972 to September 2019, Lequertier *et al.* divided the timing of prediction into four categories: before admission, at admission, throughout inpatient stay and unknown [10]. Among these studies, 4.1% (3/74) predicted LOS before admission, 29.7% (22/74) at admission, 37.8% (28/74) throughout inpatient stay and 32.4% (24/74) did not specify prediction time.

LOS prediction throughout inpatient stay is often treated as a dynamic task that involves handling time series data, such as sequences of vital sign measurements and lab test results. Signal processing techniques and recurrent neural network architectures are favored in this context, given their success in handling temporal sequence data [17, 11, 14]. In a recent study by Google researchers, Rajkomar *et al.* developed and ensembled three types of deep learning architectures with temporal mechanisms to predict long length of stay at admission and 24h after admission. In addition to the demographics and clinical diagnoses, they further included vital sign signals, lab tests and clinician's notes in the feature space, achieving the area under the receiver operating curve (AUROC) of 0.85-0.86 [18].

LOS prediction before and on admission are typically considered as point prediction problems where the models make a one-time, final decision and do not concern with updating the prediction. The primary reason is that the information available before and on admission are mostly static health records rather than dynamic series. To the best of our knowledge, there has not been any prior studies that aim to predict LOS at scheduling time. In this study, we focus on prediction at the scheduling time, which could be months prior to inpatient admission. In the later chapters, we discuss the results of LOS modeling upon admission and post operation.

Chapter 2

Data Collection

2.1 Data Inclusion Criteria

This study was conducted on the retrospective patient data provided by Boston Children's Hospital, a large pediatric medical center serving an urban population. We include a total of 22,332 elective surgical cases of 18,443 patients across all hospital service lines. All patients in this study were admitted to BCH for scheduled surgeries from January 1, 2018 to March 30, 2022, and aged between 0 and 35 at the time of admission. We exclude emergency and ICU patients, and cases without any procedural codes according to Current Procedural Terminology (CPT) defined by the American Medical Association [19]. We retain a total of 17,415 patients with 20,789 corresponding visits.

2.2 Data Overview

For each surgical case, we collect demographics, diagnoses, medication history, surgical procedure and patient flow data (e.g. admission and discharge time) from the hospital's information system. Demographics and LOS outcome are summarized in 2.1. Clinical variables, including surgical procedures, diagnoses and medication history, are summarized in Table 2.2. A detailed summary of CCSR diagnoses is summarized in the appendix A.1. The training data contains patient visits admitted during January 1, 2018 and October 1, 2021, while the out-of-sample test data are the surgical inpatient visits admitted during October 2, 2021 and March 30, 2022.

Table 2.1: Summary statistics of demographics variables and LOS

	Training data (n = 19281)	Test data (n = 1777)
Demographics		
Age, median (IQR) yr	9 (3–16)	10 (3–17)
Age categories, yr, no. %		
0–1	1750 (9.1%)	159 (8.9%)
1–2	1513 (7.8%)	161 (9.1%)
2–3	1315 (6.8%)	114 (6.4%)
3–5	3014 (15.6%)	254 (143%)
6–8	1837 (9.5%)	163 (9.2%)
9–11	1775 (9.2%)	135 (7.6%)
12–14	2143 (11.1%)	200 (11.3%)
15–17	2918 (15.1%)	272 (15.3%)
≥ 18	3016 (15.6%)	319 (18.0%)
Gender female, no. %	10588 (51.0%)	887 (49.9%)
Primary language, no. %		
English	17048 (88.4%)	1575 (88.6%)
Spanish	915 (4.7%)	91 (5.1%)
Arabic	447 (2.3%)	42 (2.4%)
Unable to collect	223 (1.2%)	1 (0.1%)
Portuguese	158 (0.8%)	21 (1.9%)
Haitian Creole	53 (0.3%)	2 (0.1%)
Cape Verde Creole	53 (0.3%)	5 (0.3%)
Chinese Mandarin	47 (0.2%)	4 (0.2%)
Sign Language	46 (0.2%)	7 (0.4%)
Hebrew	42 (0.2%)	1 (0.1%)
Vietnamese	21 (0.1%)	2 (0.1%)
Chinese Cantonese	21 (0.1%)	0 (0.0%)
Other languages	102 (0.5%)	26 (1.5%)
Major region, no. %		
Local	12798 (66.4%)	1214 (68.3%)
Regional	3351(17.4%)	340 (19.1%)
National	2431 (12.6%)	156 (8.8%)
International	655 (3.4%)	66 (3.7%)
Unknown	45 (0.2%)	1 (0.1%)
Distance traveled, median (IQR) mi	30.9 (14.4–71.3)	29.5 (14.1–60.0)
Outcome		
LOS, median (IQR) d	1.3 (1.1–3.2)	1.3 (1.1–3.2)
Number of nights, no. %		
≤ 1	11437 (59.3%)	1007 (56.7%)
2	2578 (13.4%)	253 (14.2%)
3	1881 (9.8%)	205 (11.5%)
4	1191 (6.2%)	99 (5.6%)
5	573 (3.0%)	61 (3.4%)
≥ 6	2151 (11.2%)	152 (8.6%)
Post-operative LOS, median (IQR) d	1.1 (0.9–2.9)	1.1 (0.9–2.9)
Number of nights, no. %		
≤ 1	11641 (60.4%)	1029 (57.9%)
2	2476 (12.8%)	243 (13.7%)
3	1896 (9.8%)	194 (10.9%)
4	1148 (6.0%)	110 (6.2%)
5	562 (2.9%)	56 (3.2%)
≥ 6	1558 (8.1%)	144 (8.1%)

Table 2.2: Summary statistics of procedure variables

	Training data (n=19281)	Test data (n=1777)		Training data (n=19281)	Test data (n=1777)
Surgical Procedures					
CPT					
Per-case CPT count, median (IQR)	2 (1–3)	1 (1–2)			
Level-2 CPT group, no. %			Level-2 CPT group, no. %		
Anesthesia for Procedures on the Head	72 (0.4%)	9 (0.5%)	Cardiovascular Procedures	38 (0.2%)	0 (0.0%)
Chemistry Procedures	2 (0.0%)	0 (0.0%)	Diagnostic Radiology (Diagnostic Imaging) Procedures	314 (1.6%)	10 (0.6%)
Diagnostic Ultrasound Procedures	85 (0.4%)	14 (0.8%)	Dialysis Services and Procedures	3 (0.0%)	0 (0.0%)
Evaluation and Management Services	12 (0.1%)	0 (0.0%)	Hydration, Therapeutic, Prophylactic, Diagnostic	467 (2.4%)	4 (0.2%)
Injections and Infusions, and Chemotherapy and Other Highly Complex Drug or			Intersex Surgery	13 (0.1%)	0 (0.0%)
Highly Complex Biologic Agent Administration			Introduction, Revision, and/or Removal	9 (0.0%)	0 (0.0%)
Medicine Services and Procedures	534 (2.8%)	32 (1.8%)	Neurology and Neuromuscular Procedures	17 (0.1%)	0 (0.0%)
Nuclear Medicine Procedures	12 (0.1%)	0 (0.0%)	Ophthalmology Services and Procedures	24 (0.1%)	0 (0.0%)
Physical Medicine and Rehabilitation Evaluations	3 (0.0%)	2 (0.1%)	Radiologic Guidance	11 (0.1%)	5 (0.3%)
Reproductive Medicine Procedures	1 (0.0%)	0 (0.0%)	Special Otorhinolaryngologic Services and Procedures	548 (2.8%)	39 (2.2%)
Surgical Pathology Procedures	1 (0.0%)	0 (0.0%)	Surgical Procedures on the Auditory System	1543 (8.0%)	117 (6.6%)
Surgical Procedures on the Cardiovascular System	500 (2.6%)	15 (0.8%)	Surgical Procedures on the Digestive System	8570 (44.4%)	655 (36.9%)
Surgical Procedures on the Endocrine System	239 (1.2%)	31 (1.7%)	Surgical Procedures on the Eye and Ocular Adnexa	217 (1.1%)	9 (0.5%)
Surgical Procedures on the Female Genital System	375 (1.9%)	53 (3.0%)	Surgical Procedures on the Hemic and Lymphatic Systems	200 (1.0%)	18 (1.0%)
Surgical Procedures on the Integumentary System	1984 (10.3%)	144 (8.1%)	Surgical Procedures on the Male Genital System	638 (3.3%)	63 (3.5%)
Surgical Procedures on the Mediastinum and Diaphragm	121 (0.6%)	5 (0.3%)	Surgical Procedures on the Musculoskeletal System	11181 (58.0%)	1131 (63.6%)
Surgical Procedures on the Nervous System	2736 (14.2%)	236 (13.3%)	Surgical Procedures on the Respiratory System	5874 (30.5%)	522 (29.4%)
Surgical Procedures on the Urinary System	2140 (11.1%)	272 (15.3%)			
Medication History					
Per-case level-1 medications categories count, no. %					
0	11616 (60.2%)	890 (50.0%)			
1	3352 (17.4%)	323 (18.2%)			
2	1907 (9.9%)	169 (9.5%)			
3	1022 (5.3%)	132 (7.4%)			
≥ 4	1384 (7.2%)	263 (14.8%)			
Level-1 medication category, no. %					
Alternative medicines	783 (4.1%)	112 (6.3%)	Immunologic agents	113 (0.6%)	35 (2.0%)
Anti-infectives	239 (1.2%)	61 (3.4%)	Metabolic agents	135 (0.7%)	21 (1.2%)
Antineoplastics	79 (0.4%)	18 (1.0%)	Miscellaneous agents	126 (0.7%)	19 (1.1%)
Cardiovascular agents	748 (3.9%)	128 (7.2%)	Nutritional products	2519 (13.1%)	349 (19.6%)
Central nervous system agents	2327 (12.1%)	342 (19.2%)	Psychotherapeutic agents	660 (3.4%)	101 (5.7%)
Coagulation modifiers	173 (0.9%)	34 (1.9%)	Respiratory agents	3055 (15.8%)	365 (20.0%)
Gastrointestinal agents	1690 (8.8%)	280 (15.8%)	Topical agents	1210 (6.3%)	175 (9.8%)
Genitourinary tract agents	127 (0.7%)	26 (1.5%)	Others	2438 (12.6%)	314 (17.7%)
Hormones/hormone modifiers	914 (4.7%)	124 (7.0%)			
Diagnoses					
Body System, no. %			Body System, no. %		
Cardiovascular	4598 (23.8%)	438 (24.6%)	Neurologic	5365 (27.8%)	503 (28.3%)
Digestive	5123 (26.6%)	498 (28.0%)	Nutrition	2122 (11.0%)	247 (13.9%)
Endocrine	1897 (9.8%)	196 (11.0%)	Optic	1801 (9.3%)	170 (9.6%)
Genetic	1103 (5.7%)	101 (5.7%)	Oral	1305 (6.8%)	92 (5.2%)
Hematologic	1719 (8.9%)	186 (10.5%)	Otic	2892 (15.0%)	240 (13.5%)
Immunologic	1132 (5.9%)	126 (7.1%)	Renal	1895 (9.8%)	206 (11.6%)
Infectious	10 (0.1%)	0 (0.0%)	Respiratory	7129 (37.0%)	503 (28.3%)
Mental	5886 (30.5%)	538 (30.3%)	Skin	933 (4.8%)	80 (4.5%)
Metabolic	1254 (6.5%)	149 (8.4%)	Uncategorized	2309 (12.0%)	237 (13.3%)
Musculoskeletal	7072 (36.7%)	608 (34.2%)	Urogenital	2864 (14.9%)	308 (17.3%)
Neoplasm	1149 (6.0%)	116 (6.5%)			
CCSR					
Per-case CCSR diagnoses count, median (IQR)	3 (1–8)	3 (1–9)			

2.3 Outcome Definition

Hospital length of stay quantifies the time a patient spent in the hospital from admission to discharge. In this study, we adopt a categorical representation of the LOS outcome by transforming it to the number of hospital nights, which is the time difference between admission and discharge measured in days. Moreover, we group the surgical cases with less than two hospital nights into the category of ≤ 1 for a short stay, and cases with six or more nights into ≥ 6 representing a long hospital stay. The full outcome categories are $\{\leq 1, 2, 3, 4, 5, \geq 6\}$. Similarly, post-operative LOS is defined as the time difference between surgery end time and discharge time measured in days, classified into one of $\{\leq 1, 2, 3, 4, 5, \geq 6\}$.

In particular, we do not represent LOS as a continuous variable for the following two reasons. First, the scheduling team in the hospital does not require such a high-resolution prediction. LOS estimation by days is sufficient to assist them in surgical scheduling. Second, there is randomness in the patient time course, whose factors could not be captured in the EHRs and thus the models. After patients are notified of discharge, typically before noon, they may not leave the hospital instantaneously. Some patients might stay for an extra few hours until their family is available to pick them up.

Chapter 3

Exploratory Data Analysis

Before delving into predictive modeling, we conduct exploratory data analysis to better understand the dataset by visualizing each available feature and its potential relation to the outcome. In this chapter, we first discuss the outcome distribution, and then analyze the features in demographics, surgical procedure, diagnosis and medication history. Finally, we explore the patient care class and temporal features (e.g. admission time, operative length etc.), which are available on or after inpatient admission.

3.1 LOS Outcome

Fig 3.1 shows the LOS outcome distribution of the full population. Over 59% of the surgical inpatient visits have a short length of stay for no more than 1 day, and 8.4% of the cases have a long LOS for more than 5 days. The majority class ($\leq$ 1d) is 20 times as many as the size of the smallest class (= 5d), which indicates a moderate class imbalance problem.

Fig 3.2 shows the outcome class distribution stratified by time period. The relative class size is consistent across all time periods–short LOS is the majority class that occurs in over 55% of the cases, and class LOS = 5 nights is the minority class that happens less than 4% among the full population. Inpatient visits with long LOS (6 nights or more) take up 8–10% during 2018–2021.

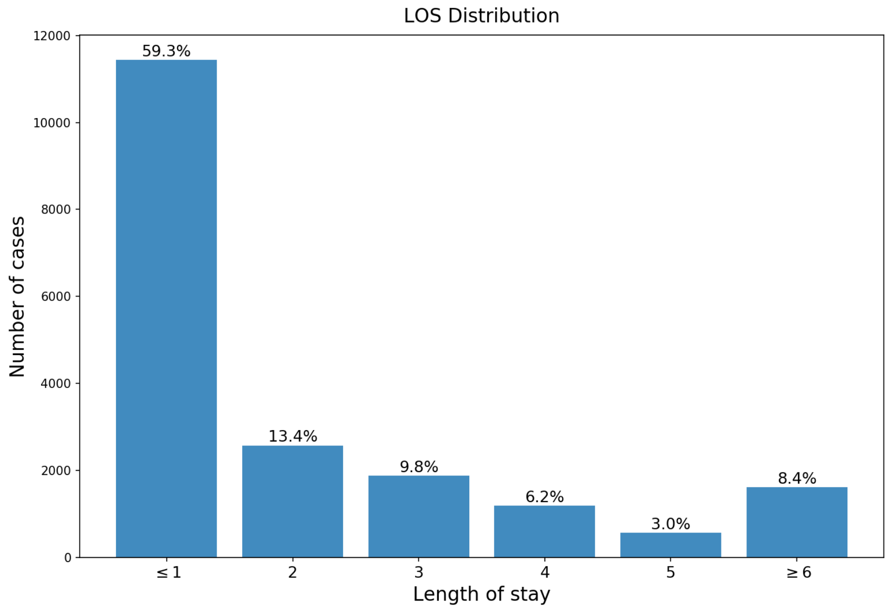


Figure 3.1: LOS outcome distribution on the entire population. y-axis is the count of surgical cases. Above each bar is the percentage of the class size relative to the full population size.

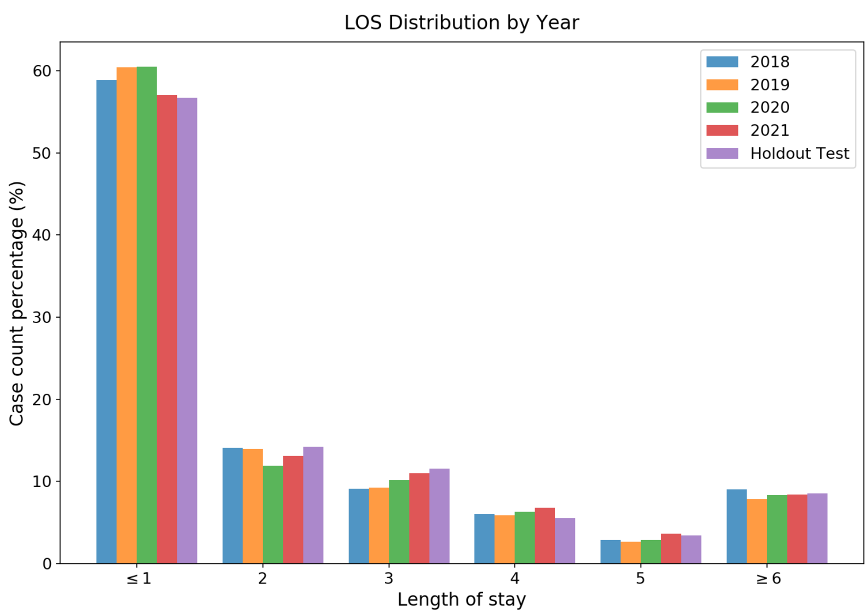


Figure 3.2: Stratified LOS outcome distribution. y-axis is the percentage of the class size relative to each stratified category sample size.

3.2 Demographics

3.2.1 Gender

According to Fig 3.3, among all surgical inpatient visits, 49.05% are male and 50.95% are female. Over the six outcome classes, male and female patients differs in case count at LOS = 3d, where more female patients are hospitalized for 3 nights than male patients. Among other classes, there are approximately the same number of male and female inpatient visits, which implies that gender alone is not a strong indicator of the LOS outcome.

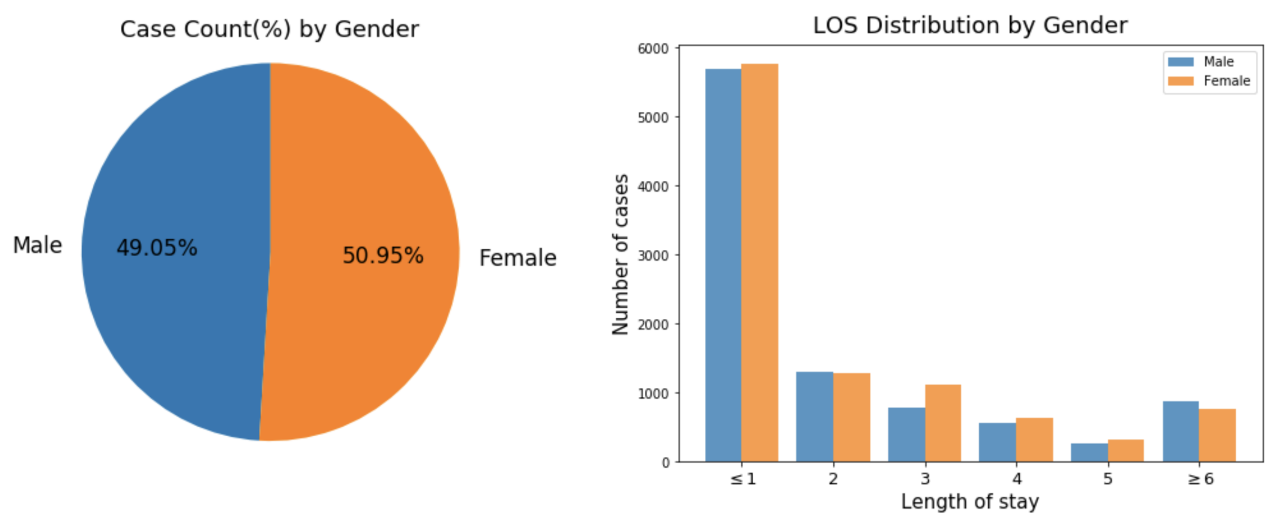


Figure 3.3: Gender and LOS outcome. Left plot is the relative percentage of male and female patients. Right plot is the LOS outcome distribution stratified by gender.

3.2.2 Age

To understand the relation between age and LOS, we stratify the patient population by age into the following groups: 0–3 mos, 3–6 mos, 6 mos–1 yr, 1–2 yrs, 2–3 yrs, 3–6 yrs, 6–9 yrs, 9–12 yrs, 12–15 yrs, 15–18 yrs and > 18 yrs according to the developmental stages defined by CDC [20]. From Fig 3.4, the majority of cases have a short LOS in each age group. Patients older than 6 years is more likely to stay longer, given the slightly heavier tail at the upper LOS ranges. However, age alone does not show obvious correlation with LOS outcome.

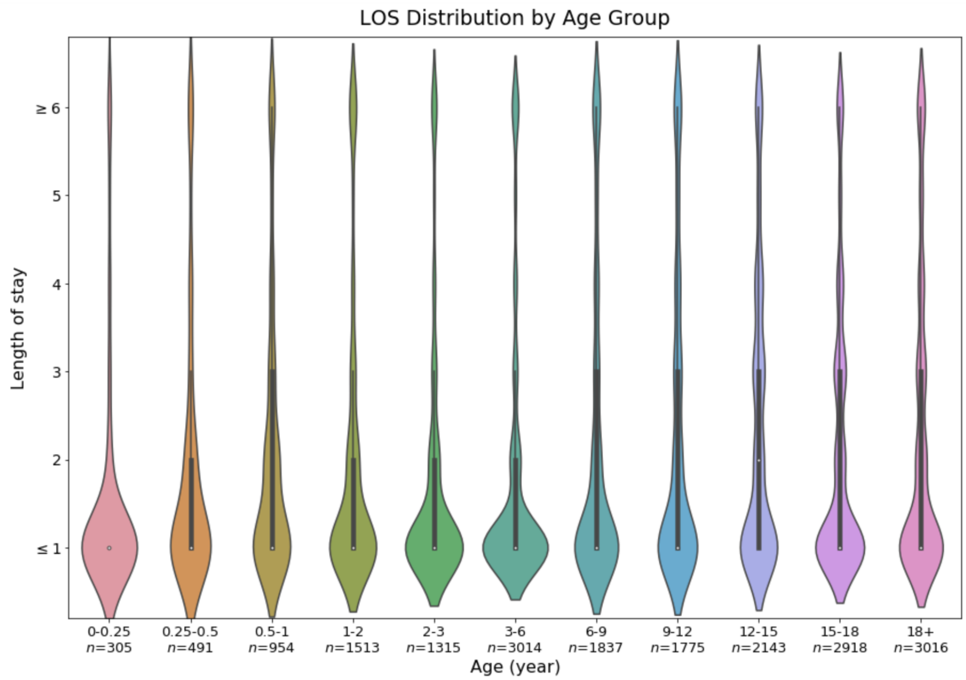


Figure 3.4: Age and LOS outcome violin plot. Each "violin" shows the outcome distribution within the corresponding age group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome.

3.2.3 Language

Top 15 Most Frequent Languages

We further explore the relation between patient's mother language, interpreter need and LOS outcome. Fig 3.5 shows the LOS outcome distribution over the top 15 most frequent language groups. Patients who speak Arabic, Haitian Creole, Hebrew and Vietnamese more often has longer LOS. Therefore, whether a patient speaks any of these languages is potentially indicative of LOS outcome. On the contrary, other non-English speakers mostly stay for 1 day or less.

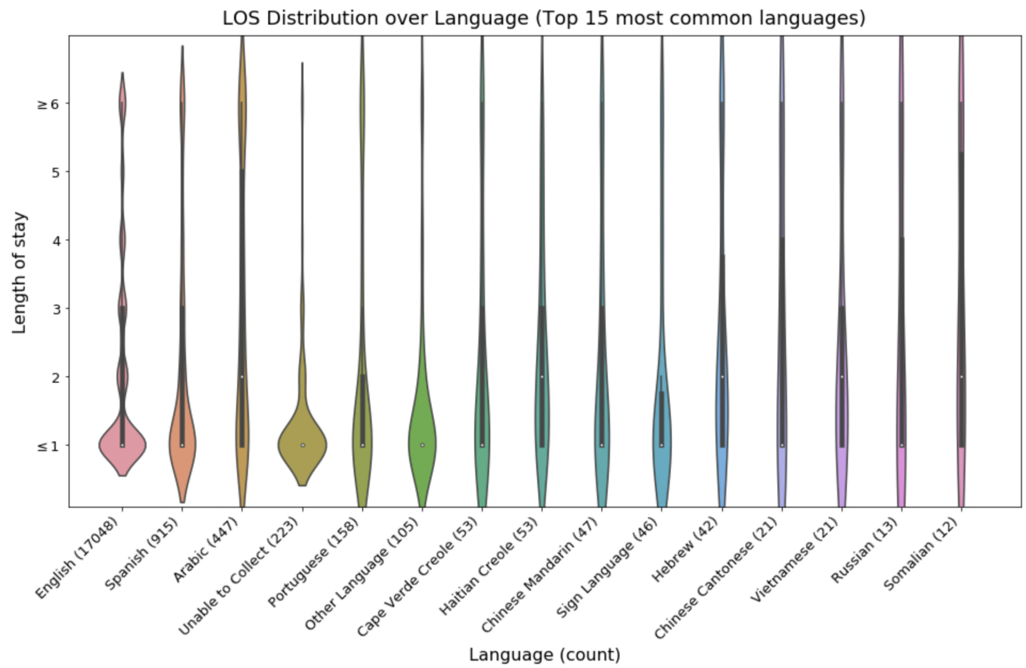


Figure 3.5: Mother language and LOS outcome violin plot. Each "violin" shows the outcome distribution within the corresponding language group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome. The number in parenthesis is the count of surgical cases within each language group.

Interpreter Need

In Fig 3.6, we visualize the impact of interpreter need on LOS outcome for the top 15 most frequent language groups described above. Patients who speak Spanish, Arabic, Portuguese, Cape Verde Creole, Haitian Creole and Vietnamese have almost identical LOS outcome distribution no matter the interpreter need. Other language groups, such as Mandarin, Cantonese, Russian and Somalian, exhibit different outcome distribution patterns. However, there are fewer number of cases in either sub-group based on interpreter need, so we cannot draw a statistically significant conclusion on the impact of interpreter need on LOS outcome. Overall, interpreter need is unlikely to be predictive of the LOS outcome.

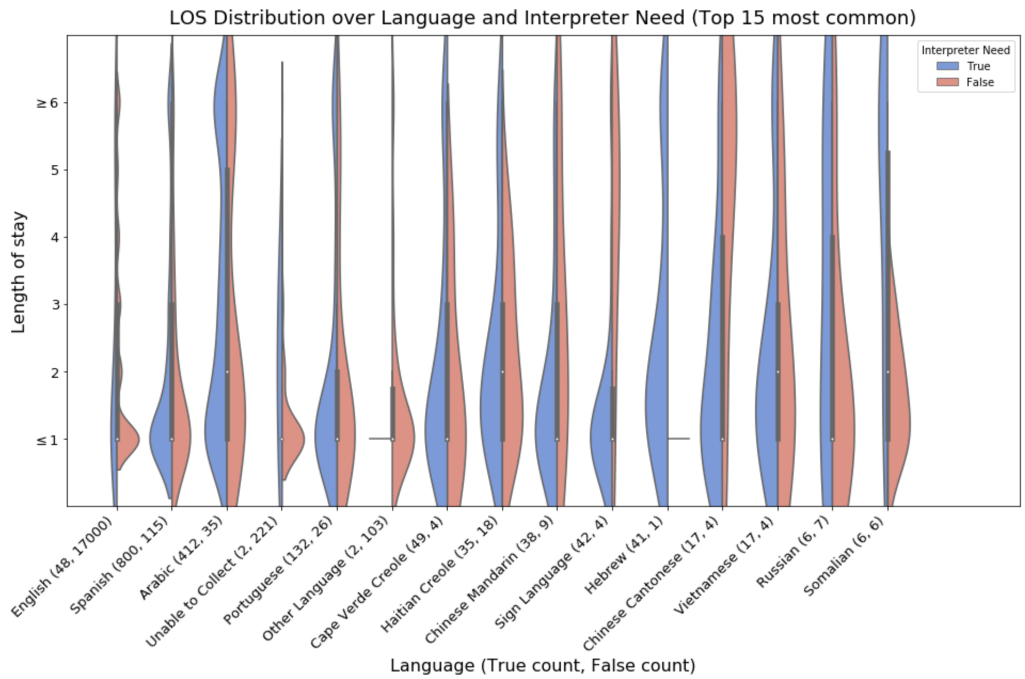


Figure 3.6: Interpreter need by language and LOS outcome violin plot. Each compound "violin" shows the outcome distribution within the corresponding language group.The blue half represents inpatient visits that require an interpreter, while the pink half those who do not. The width is proportional to the sample size of each group. The white dot is the outcome median of the language group. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome for the corresponding language group. The numbers in parenthesis are the count of surgical cases within each language group, left number for patients who need interpreter and right number for those who do not. A line segment means all data has a single LOS outcome.

3.2.4 Residence Location

Major Region

Fig 3.7 shows the LOS outcome distribution over five types of patient's major region of residence. Local patients reside in Massachusetts; regional patients are from outside Massachusetts but within the New England area; national patients live outside the New England area but within the United States; international patients come from outside the United States. The median outcome class for national and international patients is 2 days, and more frequently stay for more than 5 days, compared to local and regional patients. Such difference implies that major region might have some predictive power of the LOS outcome.

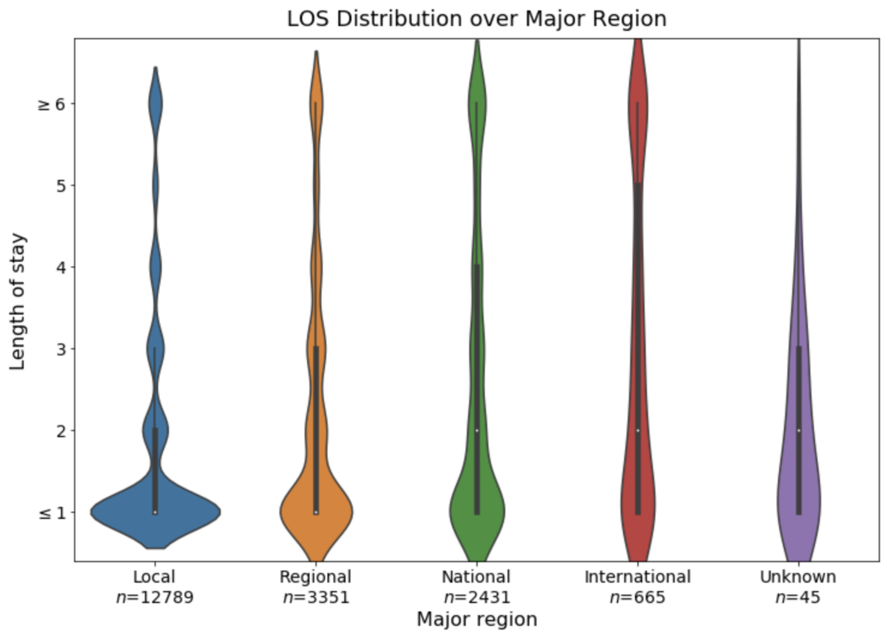


Figure 3.7: Major region of residence and LOS outcome violin plot. Each "violin" shows the outcome distribution within the corresponding region group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome.

Distance to Hospital

The actual distance between patient's residence and the hospital is calculated by zip code, and provides a higher resolution to understand the impact of residence location on LOS outcome. Based on Fig 3.8, LOS outcome have similar distribution with the majority of cases centering around LOS $\leq$ 1d when the distance is lower than 66.3 miles. Patients who live further from the hospital tend to have longer LOS, given a slightly heavier tail at the upper outcome range. It matches our observation on major region. Moreover, residence location characterized in distance does not necessarily provide additional information to LOS prediction.

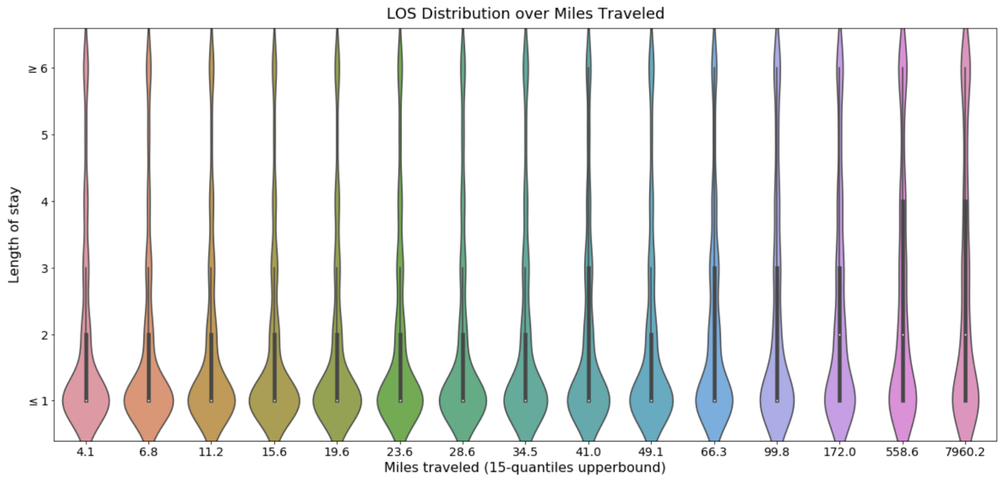


Figure 3.8: Distance to hospital and LOS outcome violin plot. Distance (measured in miles) is grouped by 15-quantiles. Each "violin" shows the outcome distribution within the corresponding distance group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome.

3.3 Surgical Procedures

Surgical procedure comes in two formats at Boston Children's Hospital: primary procedures and CPT code. Primary procedures are defined by the SurgiNet, a module of Cerner Millennium, designed specifically for surgical usage. CPT is the coding of all medical services and procedures a healthcare provider offers defined by AMA [21]. Each surgical inpatient visit has exactly one primary procedure, but could have multiple CPT codes when the patient needs auxiliary procedures or undergoes multiple procedures from different medical specialties. To the best of our knowledge, there is no clinician-endorsed, one-to-one mapping from SurgiNet primary procedures to CPT codes. Therefore, in this section, we explore the two formats of surgical procedure coding independently.

3.3.1 Primary Procedure

In the training data, there is a total of 1,273 primary procedures, among which the most common procedure, tonsillectomy with adenoidectomy, covers 9.4% (1809 / 19281) cases.

However, there are 333 rare procedures (e.g. hemipelvectomy, toe transfer, etc.) showed up only once out of 19,281 inpatient visits. To ensure statistical significance, we visualize the LOS (in continuous format) distribution of the top 20 most common primary procedures, each of which has at least 135 corresponding inpatient visits.

Top 20 most common primary procedures

Fig 3.9 demonstrates that spine fusion and hip PAO patients had considerable longer LOS than patients who underwent other procedures. With median and the lower quartile both over 3 days, the majority of these inpatients has an LOS of at least 3 days. However, all three distributions have a large IQR range, which indicates a high variance in outcome.

In comparison, procedures such as breast reduction and tonsillotomy with adenoidectomy have much narrower distribution where the majority of the inpatient visits stayed for 1 day. It is important to note that, despite their narrow IQR range, there still exists quite a few outliers that stayed for two or more days.

Overall, the characteristics of the LOS distribution of each procedure cohort is distinct and imply different level of surgical complexity. Therefore, primary procedure is likely to provide important information to predicting LOS outcome.

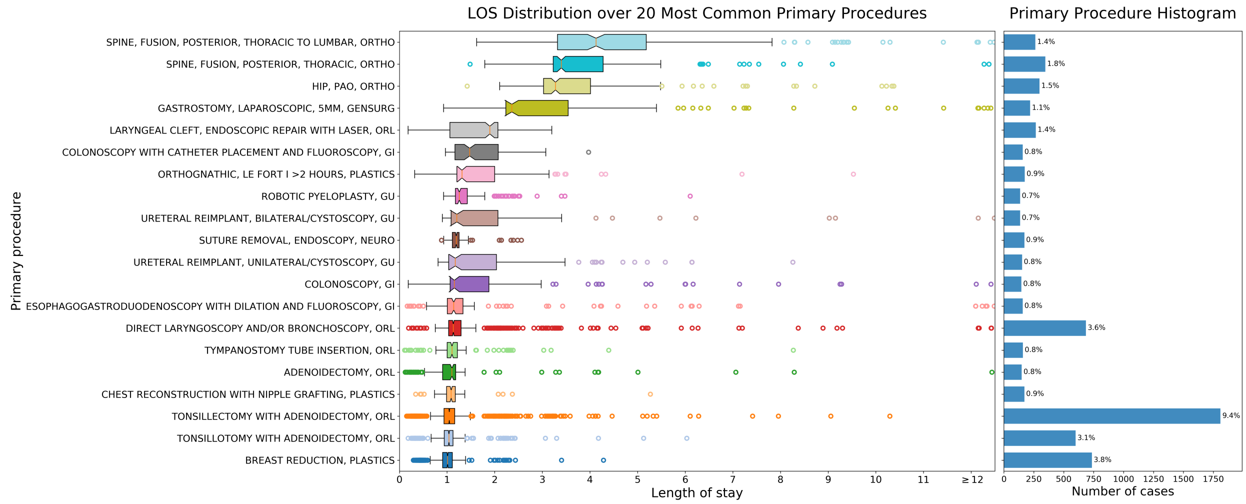


Figure 3.9: Top 20 most common primary procedures and LOS distribution. In the left plot, x-axis is the LOS measured in hours and converted to the unit of day. Cases with LOS > 12 days are represented as a random float between 12 and 12.5. y-axis for both plots represents the top 20 most common primary procedures sorted according to the cohort's LOS median in descending order from top to bottom. The plot to the right the histogram of the corresponding primary procedures labeled with their frequency in percentage.

Procedure cohort LOS median and LOS outcome class

We further examine the relation between the procedure-wise LOS median and the outcome class. To obtain a more clear picture, we round each cohort median to the nearest integer and visualize its outcome distribution in a violin plot (see Fig 3.10). We restrict the cohort LOS median to a maximum of 15 days, and group the cohorts with outlying LOS median ($\geq$ 15d) into a separate category.

The violin plot shows an obvious pattern that surgical cohorts with higher median LOS distribute more frequently around the upper range of the outcome class. In group 1 where the cohort median LOS is 1 day, almost all cases fall under the outcome class of LOS $\leq$ 1d. In contrast, from group 6 and beyond, the majority of the inpatient visits fall under outcome class of LOS $\geq$ 6d. However, there still exist a few outliers with a short LOS. This could happen when the surgeons discover during the procedure that a patient's condition is so complex that they have to terminate the surgical operation and consequently discharge the patient early.

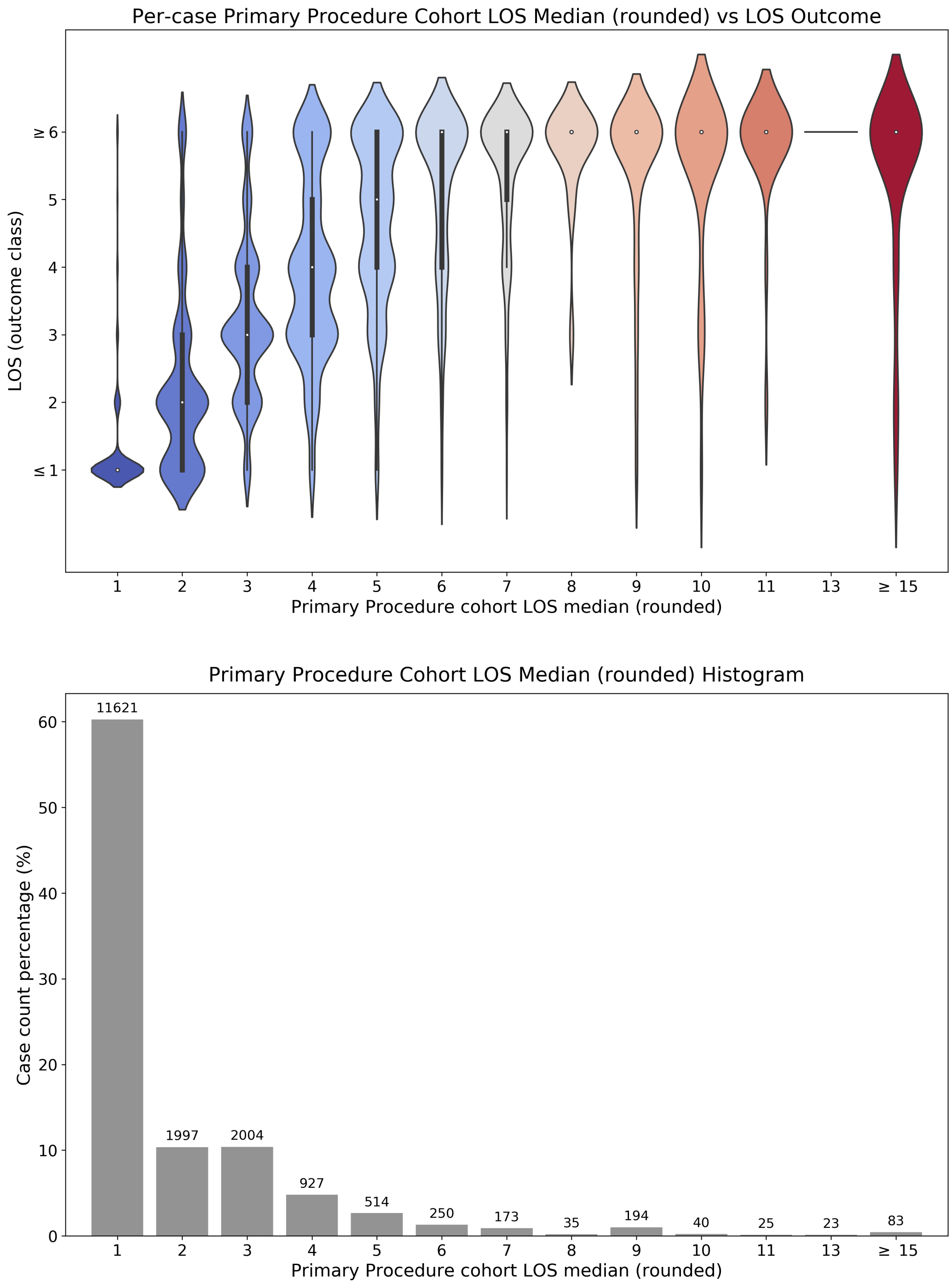


Figure 3.10: Primary procedure cohort LOS median (rounded) and LOS outcome violin plot and histogram. We filter out the cases whose primary procedure appeared less than 5 times in the training set, retaining a total of 17,886 cases for generating these plots. In the upper plot, all "violins" are scaled to the same width. The bottom plot is the histogram of each LOS median group. y-axis is the sample size percentage of each group, labeled with the actual number of surgical cases.

From group 2 to 5, the cohort LOS median does not show a clear boundaries among the outcome classes from 2 to 5. For example, even though group 2 patients are most likely to fall under outcome class of LOS $= 2$d, there is still a considerable number of visits that stayed for exactly 1 day and 3 days. Similar pattern exists in group 3, 4 and 5. Therefore, cohort LOS median is indicative of very short (≤ 1d) and very long length of stay (≥ 6d), but could not clearly distinguish among the outcome classes in the middle range (2–5d).

3.3.2 CPT Code

From Fig 3.11, 48.6% of the inpatient cases have exactly one CPT code, while 51.4% cases have two or more CPT codes. In Fig 3.12, we explore the connection between the number of CPT codes per surgical case and the LOS outcome. The heavier tail at outcome class LOS ≥ 6d indicates that the surgical cases with more than five CPT codes tend to be more complex. However, the majority of the cases still fall under the class for LOS ≤ 1d. Thus, the count of CPT codes per case is not very indicative of the LOS outcome.

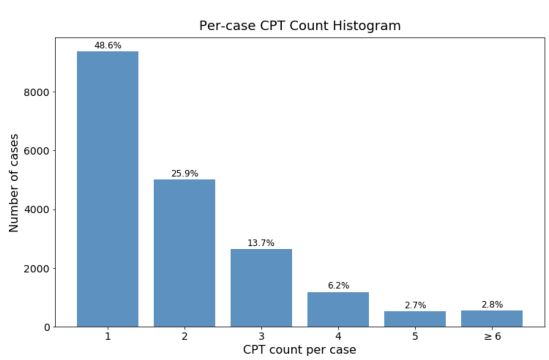


Figure 3.11: Per-case CPT count histogram. $x-$axis is the number of CPT codes per surgical case. $y-$axis is the number of surgical cases. Each bar is labeled with the percentage with respect to the full training set.

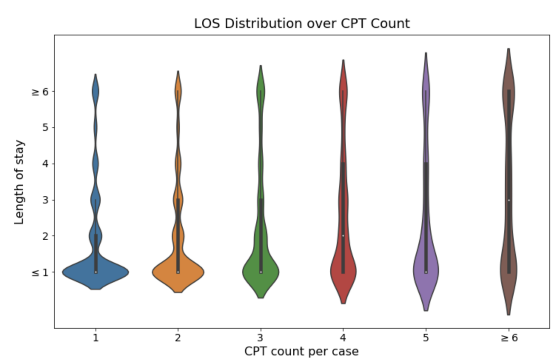


Figure 3.12: Per-case CPT count and LOS outcome. $x-$axis is the number of CPT codes per surgical case. $y-$axis is the LOS outcome class. Each "violin" shows the outcome distribution within the corresponding CPT count group. The width is proportional to the sample size of each group.

3.4 Diagnoses and Historical Conditions

Diagnosis of patients' current and chronic conditions captures their overall health status. In this section, we look into how the diagnosis information relates to the hospital length of stay.

At Boston Children's Hospital, the diagnosis is originally determined via the International Classification of Diseases, 10^{th} Revision Clinical Modification (ICD-10-CM), which is a comprehensive coding system for all medical diseases and conditions. There are over 70,000 ICD-10 diagnosis codes, but not all of them are clinically meaningful. Thus, we use the Clinical Classifications Software Refined (CCSR) that aggregates the ICD-10 codes into over 530 clinically meaningful categories across 21 body systems [22]. In this study, we adopt CCSR grouping and the body system category as the primary source of diagnosis information associated with each surgical case.

3.4.1 Body System Diagnosis

There are 21 body systems category related to each diagnosis. Fig 3.13 and Table 2.2 shows their relative frequency among the 19,281 cases in the training set. Respiratory, musculoskeletal and neurologic diagnosis are the most frequent categories, which occurred in 12.0%, 11.9% and 9.0% surgical cases respectively. On the contrary, infectious is the least common category, showing up in only ten cases.

It is important to note that each surgical case could have multiple body system diagnoses. The histogram in Fig 3.13 shows the number of cases grouped by the count of their body system categories. 12.7% cases do not have any chronic conditions, while 59.8% cases have more than one body system diagnoses. Patients without any chronic conditions are more likely to stay for one or less night in the hospital, while patients with ten or more body system diagnoses have a higher chance of staying for more than 5 nights. Patients with two body system diagnoses tend to have a long LOS more often than those with a single diagnosis (number of cases under class LOS $\geq$ 6 in group 2 is almost identical to that in group 1). In other groups, count of body system diagnosis does not indicate a strong difference in the outcome distribution.

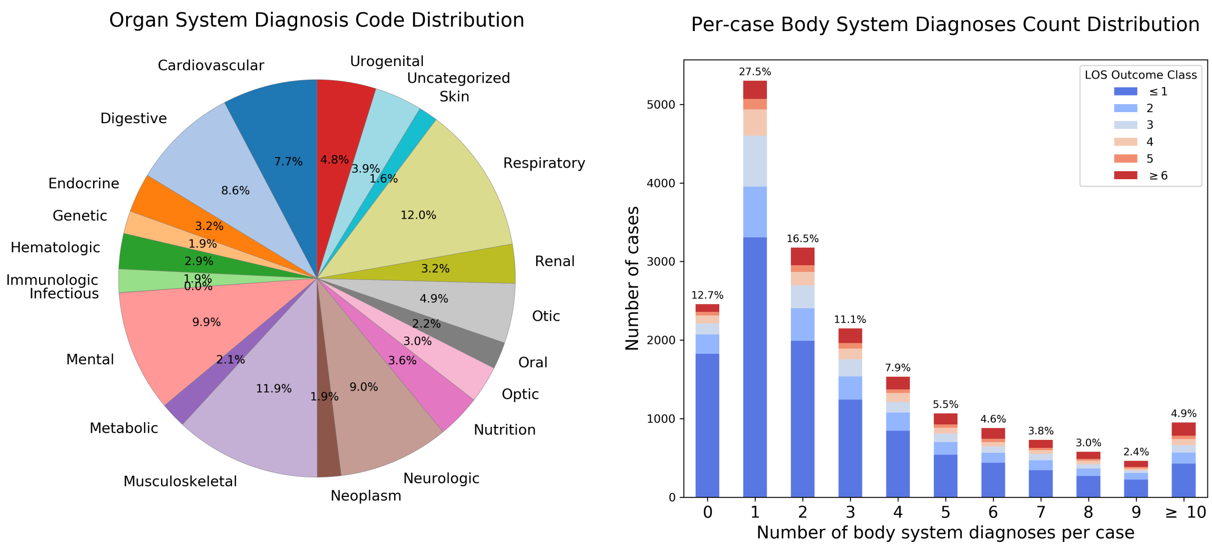


Figure 3.13: Body system diagnosis distribution (left) and count of body system diagnosis per case (right). The left plot shows the relative percentage of each body system diagnosis among the training set. The right plot is the number of body system diagnoses per patient encounter. $x-$axis is number of diagnoses per surgical case. $y-$axis is the number of surgical cases admitted during each hour of day. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each diagnosis count group.

Fig 3.14 visualizes the LOS distribution for each body system diagnosis category sorted by median in descending order from top to bottom. Patients with neoplasm diagnosis tend to have longer LOS, while otic and oral patients tend to have shorter LOS. All categories have a wide IQR range, which indicates that the body system diagnosis does not have a clear relation to LOS. In addition, since each patient could have multiple body system diagnoses, there exists mixed effects that is not untangled in the plot.

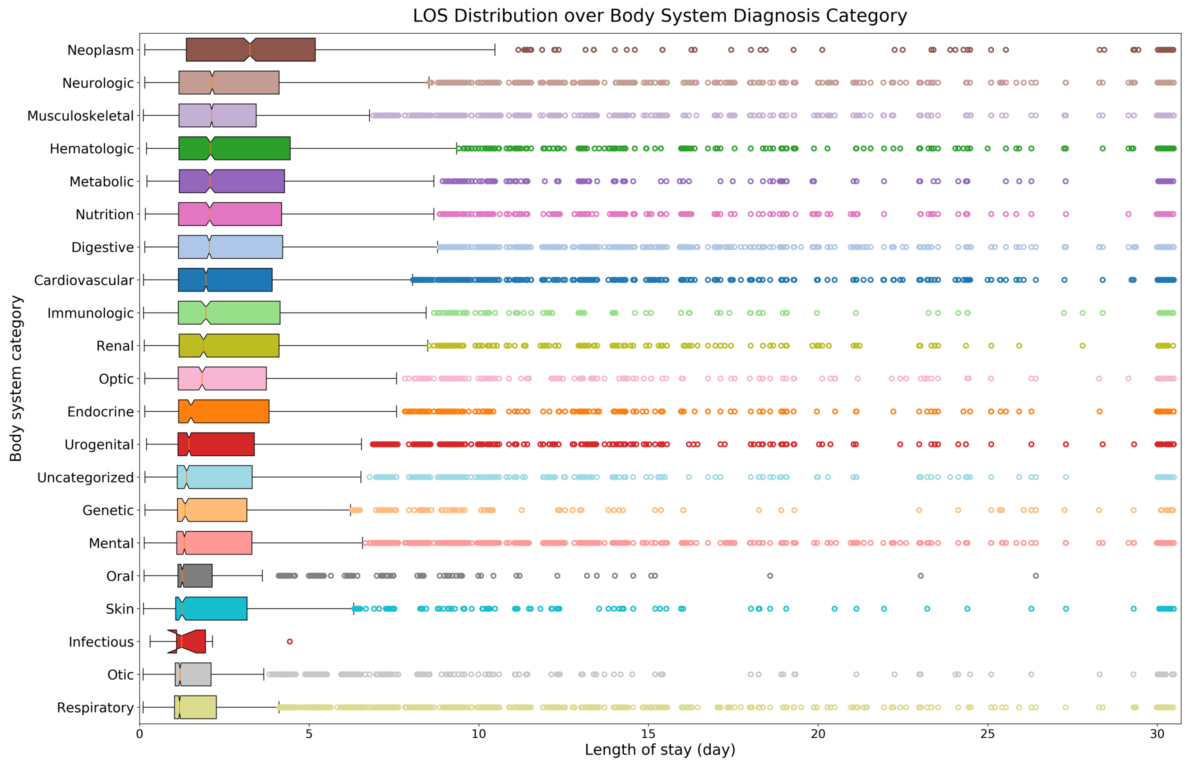


Figure 3.14: Body system diagnosis and LOS. Left plot is the frequency of each body system diagnosis in the training data. Right plot is the histogram of the body system diagnoses count per surgical case. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each body system diagnosis group.

3.4.2 CCSR Diagnosis

CCSR diagnosis codes specify the disease and chronic conditions to a more granular degree. Similar to body system diagnosis, each patient could have multiple chronic conditions, which might impact their hospital LOS. The training data covers a total of 719 CCSR diagnosis codes, wherein the median number of CCSRs per surgical case is 3 and the IQR range is $[1, 8]$. Fig 3.15 shows the histogram of per-case count of CCSR diagnosis codes in the training data.

In Fig 3.16, we explore the relation between LOS and the count of CCSR diagnosis per surgical case. From the violin plot, we see that patients tend to have longer LOS if they are diagnosed with more than seven conditions, which is evident in larger median values and higher IQR ranges.

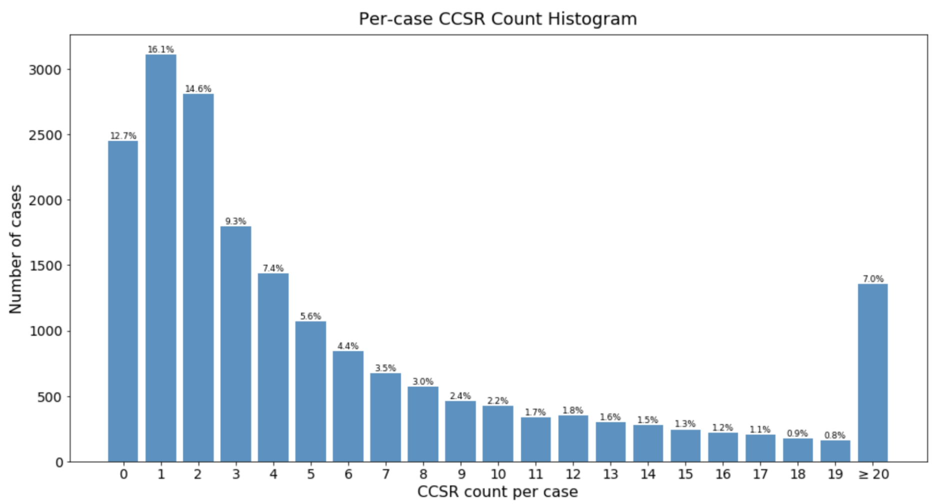


Figure 3.15: Per-case CCSR count histogram. $x-$axis is the number of CCSR diagnosis codes per surgical case. $y-$ axis is the number of surgical cases of each CCSR count group. Each bar is labeled on top with the percentage relative to the full training set.

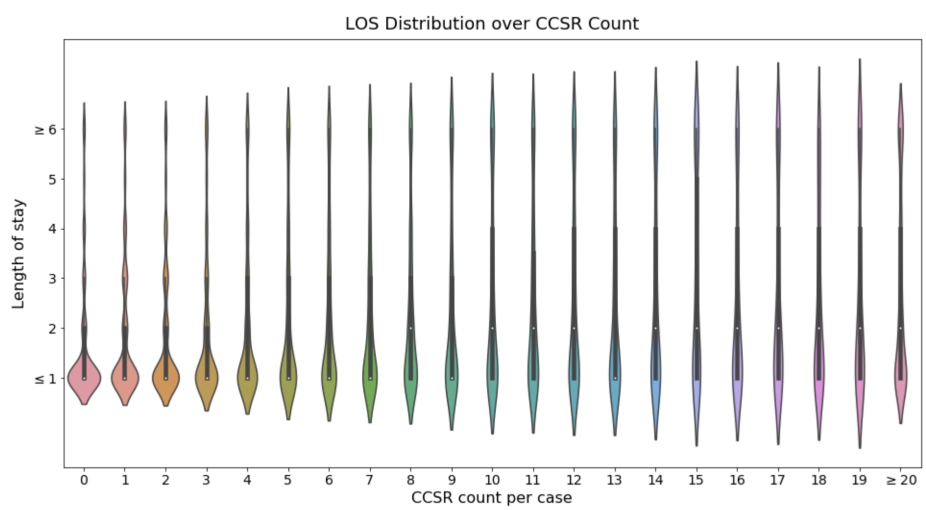


Figure 3.16: Per-case CCSR count and LOS outcome violin plot. Each "violin" shows the outcome distribution within the corresponding CCSR count group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome.

From Table A.1 in the appendix, we see that neurodevelopmental disorder, sleep apnea and hearing loss are all very common conditions, occurring in 17.3%, 16.6% and 13.5% respectivelt in the training data. However, they imply different levels of complexity, so the presence of different conditions may have different impacts on the actual LOS.

In Fig 3.17, we visualize the LOS distribution over the 20 most common CCSR diagnosis cohorts. Conditions, such as scoliosis, imply longer LOS, given that they have larger median and a higher IQR in LOS, but patients with other conditions like tonsils, sleep apnea and hearing loss are mostly hospitalized for one night. Overall, the majority of these CCSR conditions have a wide IQR range, which indicates that CCSR codes alone is unlikely to distinguish among all outcome classes.

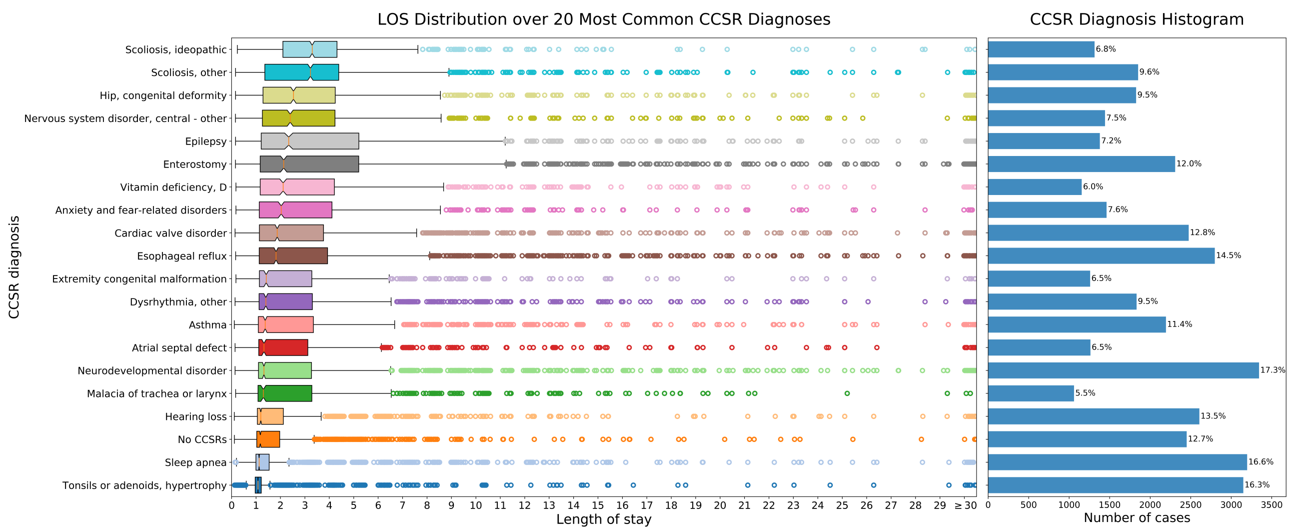


Figure 3.17: 20 most common CCSR diagnoses and their LOS distribution. In the left plot, x-axis is the LOS measured in hours and converted to the unit of day. Cases with LOS $>$ 30 days are represented as a random float between 30 and 30.5. y-axis for both plots represents the top 20 most common CCSR diagnoses sorted according to the cohort's LOS median in descending order from top to bottom. The plot to the right is the histogram of the corresponding CCSR diagnoses labeled with their frequency in percentage.

3.4.3 SNOMED Problem Count

Systemized Nomenclature of Medicine – Clinical Terms (SNOMED CT) is another way of encoding patient problems (diseases and conditions). SNOMED CT can be mapped to ICD-10 codes and its CORE (Clinical Observations Recording and Encoding) Problem

List Subset is derived from clinical datasets from large scale healthcare institutions. It is commonly used for coding clinical information at a summary level [23]. The dataset contains the count of the problem list for each surgical case. In Fig 3.18, we see that 25.3% of the cases do not have any comorbidity from SNOMED problem list. Note that this does not match the CCSR diagnoses count because SNOMED does not have a one-to-one mapping to CCSR codes. Moreover, it is highly non-standardized and used with varying purposes other than to indicate clinical conditions.

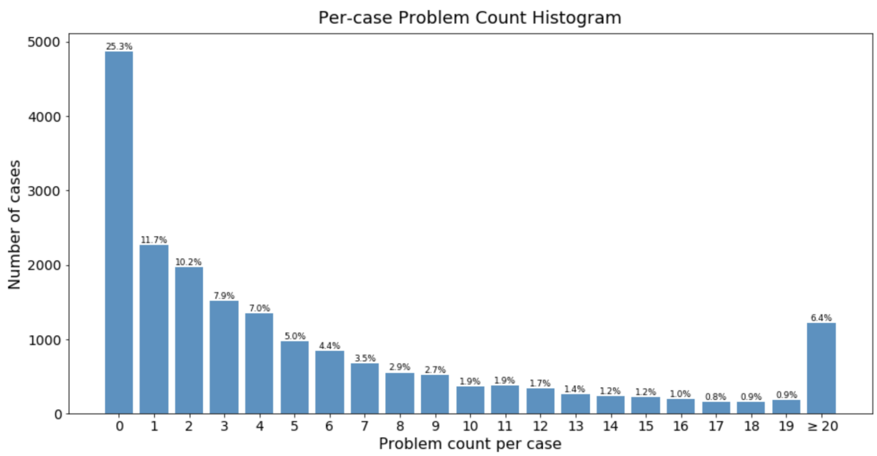


Figure 3.18: SNOMED problem count histogram. $x-$axis is the number of SNOMED problems per surgical case. $y-$axis the number of surgical cases in the training set. Each bar is labeled with the percentage of the cases within each problem count group.

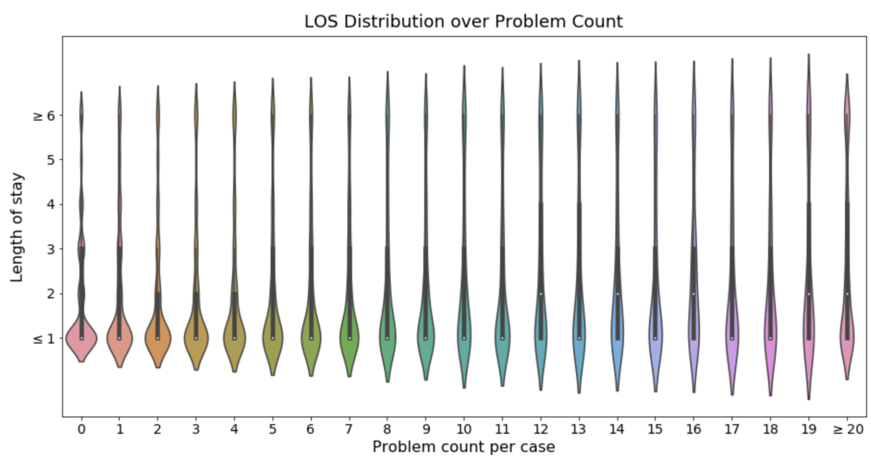


Figure 3.19: SNOMED problem count and LOS outcome. Each "violin" shows the LOS outcome distribution within the corresponding problem count group. The width is proportional to the sample size of each group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome.

3.5 Medication History

In this section, we explore patient medication history and how it might correlate with the LOS outcome. All medications in our dataset is coded via Cerner Multum, a drug, herbal and nutraceutical database. It has three levels of categorization granularity (see details in the appendix Table). In this study, we adopt the most granular possible (level-3 or below) medication group to represent the patient medication history. We filter out inactive or uncertain usage status medications and extract the full medication history of every surgical inpatient.

Fig 3.20 shows the outcome distribution of the 20 most common medication categories. Patients who had a history of using benzodiazepines and benzodiazepine anticonvulsants tend to have a longer length of stay. However, other medication types do not show a clear distinction in LOS distribution, given their highly overlapping IQR ranges and similar values of LOS median. Moreover, the IQR of all twenty medication groups are wider than that of primary procedures and CCSR diagnosis, implying a lower predictive power of LOS.

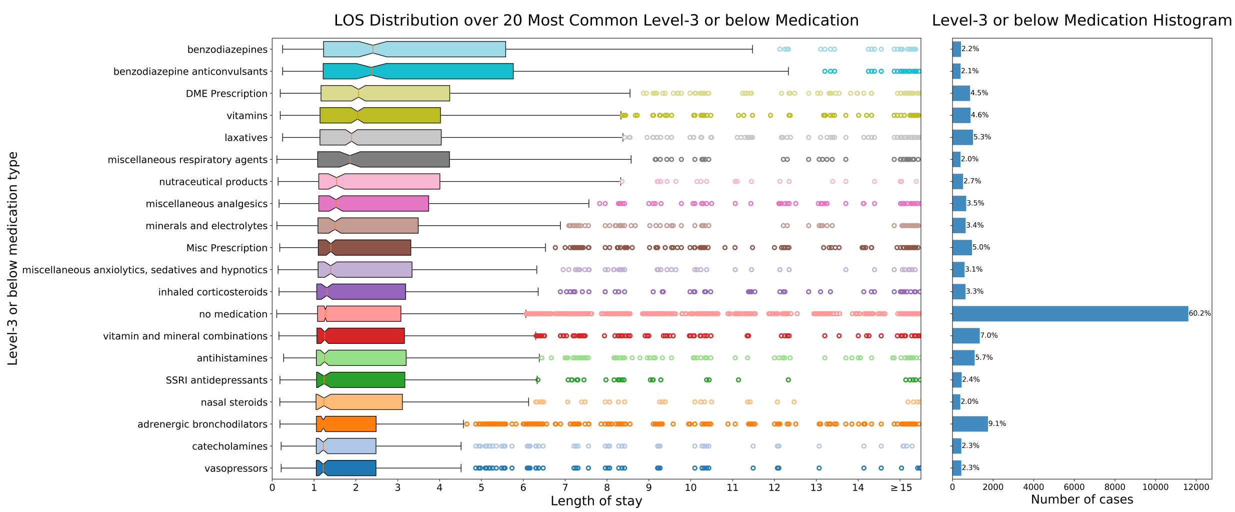
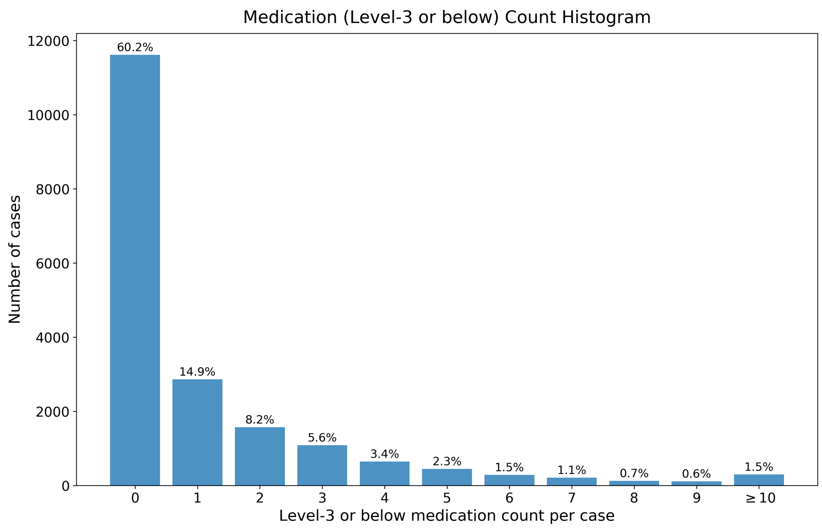


Figure 3.20: Level-3 or below medication and LOS. $y-$axis for both plots is the twenty most common medication types of level-3 or below sorted in descending order from top to bottom according to their LOS median. In the left plot, $x-$axis is the continuous value of length of stay measured in days. All cases with LOS ≥ 15 are represented as a random float around 15 for better visualization. In the right plot, $x-$axis is the number of surgical cases. Each bar is labeled with the sample size percentage respective to the full training set.

Fig 3.21 shows the histogram of the number of medication types per surgical case, as well as its association with the outcome class. Over sixty percent of the cases do not have any medication history, among which the majority of cases have an LOS of a very short LOS. As the count of past medications increases, patients are more likely to have a long LOS since they may suffer from complex chronic conditions. However, class LOS ≤ 1 still dominates among all medication count groups.



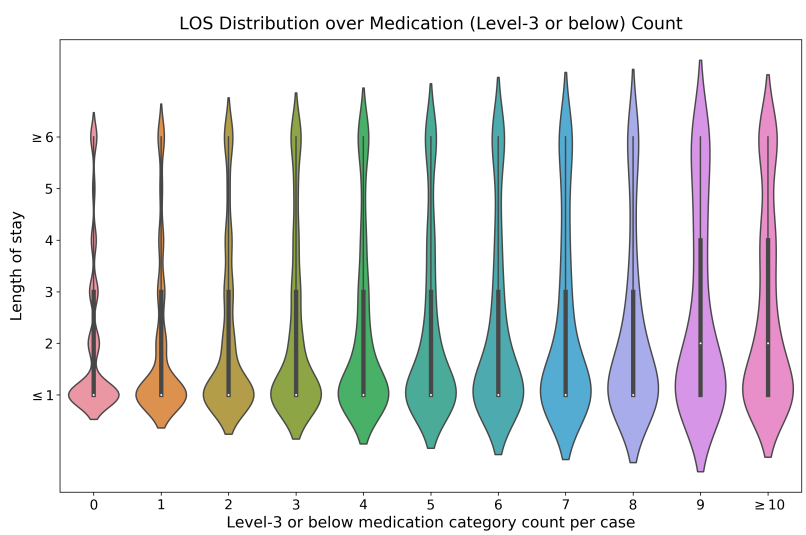


Figure 3.21: Count of level-3 or below medication types per case and LOS. Upper plot is the histogram of historical medication count per case. Bottom plot is the LOS outcome distribution over each medication count group. Each "violin" shows the outcome distribution within the corresponding medication count group. The white dot is the outcome median. The dark grey bar in the middle of each "violin" marks the IQR range of the LOS outcome. All "violins" are scaled to the same width.

3.6 Features on and after Admission

From this section onward, we discuss the features that are available on and after inpatient admission. We first discuss the inpatient care status and then focus on the temporal features throughout the patient pathway.

3.6.1 Care Class

The care status describes the level of medical and nursing care a patient would need during an inpatient visit. It has two categories: inpatient and observation. Inpatient status typically means a patient needs more complex or longer care, while observation status normally applies for patients who have an unclear need for longer care or whose conditions usually respond to less than 48 hours of care [24]. The care status is determined by physicians after careful considerations of numerous factors, including the severity of the medical conditions, patient medical history, hospital bylaws and admission policies [25].

Fig 3.22 shows the proportion of training set cases under each LOS outcome class, colored by the care class. The majority of observation patients were discharged within two nights after admission. However, inpatient care visits distribute relatively evenly among the six outcome classes. In predicting LOS, this information is likely to help distinguish very short LOS, but require other features to differentiate among the long LOS outcome classes.

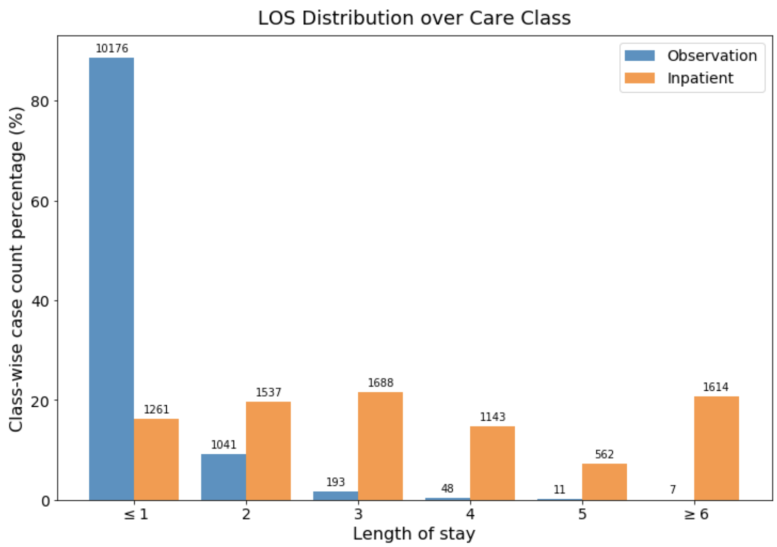


Figure 3.22: Patient care class and LOS outcome. This plot is a histogram of the LOS outcome classes stratified by patient care class. $x-$axis is the LOS outcome class. $y-$axis is the frequency of cases within each outcome group. Blue represents observational status patients. Orange represents patients under inpatient care. On top of each bar is labeled with the count of the surgical cases.

3.6.2 Admission Time

In Fig 3.23, we visualize the temporal pattern of admission by the hour of day. Over 97% of the surgical cases in the training set were admitted between 5am and 3pm. Each color in the bars represents an LOS outcome class. Based on the case distribution over the outcome, admission hour of day is not indicative of the actual LOS.

Fig 3.24 shows the surgical case distribution over admission day of week, where each color represents an outcome class. 97.9% cases were admitted during weekdays. Class "$\leq$ 1" dominates in all weekday and Saturday admissions, among which Tuesday admissions have a higher proportion of hospital stays for more than two nights. In comparison, Sunday admissions (1.1% of the training set) have a high percentage of cases with a long LOS.

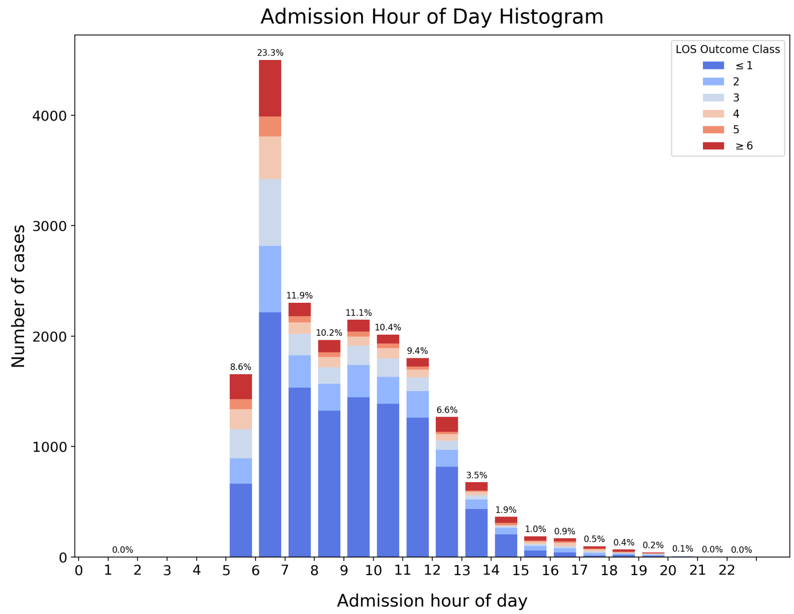


Figure 3.23: Admission time of day histogram. $x-$axis is the hour of day. $y-$axis is the number of surgical cases admitted during each hour of day. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each hour group.

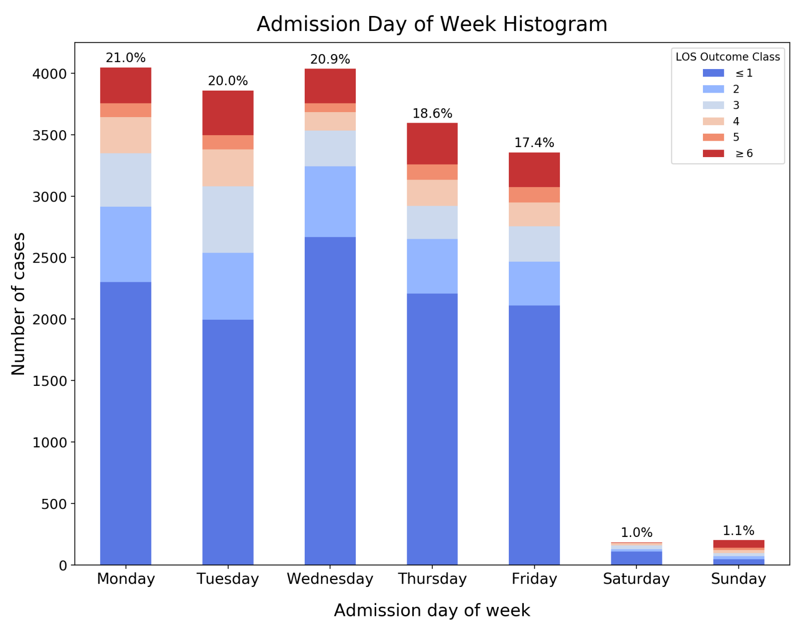


Figure 3.24: Admission day of week histogram. $x-$axis is the day of week. $y-$axis is the number of surgical cases admitted during each day of week. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each day of week group.

3.6.3 Discharge Time

Fig 3.25 shows the histogram of discharge hour of day. The red dashed line marks 11am, which is the latest time by which clinicians notify the patients whether they are eligible for discharge. Unlike admission, discharge time has a wider time range from 8am to 10pm. Multiple factors could influence the exact discharge time, including the availability of the patient's family, the variability during the discharge process, etc., but these factors are not captured by the EHRs. Therefore, in this study, we aim to predict the number of nights spent in the hospital rather than the specific hour of discharge.

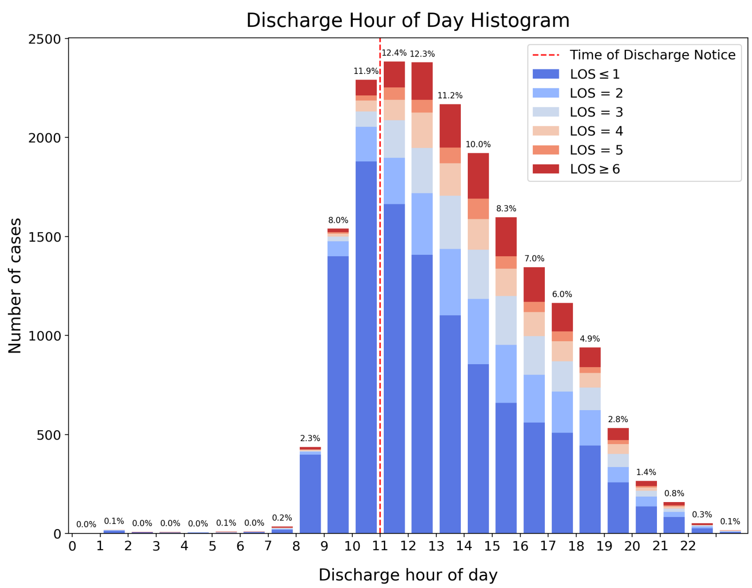


Figure 3.25: Discharge time of day histogram. $x-$axis is the hour of day. $y-$axis is the number of surgical cases discharged during each hour of day. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each hour group. The red dashed line marks 11am, which is the time of discharge notice to each patient eligible for discharge.

Fig 3.26 shows the pattern of discharge day of week. Sunday and Monday have the least number of discharges, and both have higher percentage of inpatient visits that stayed for more than two nights. Among other days of week, LOS outcome of less than two nights dominates. Tuesday has nearly no cases with an LOS of two or three nights because Saturday

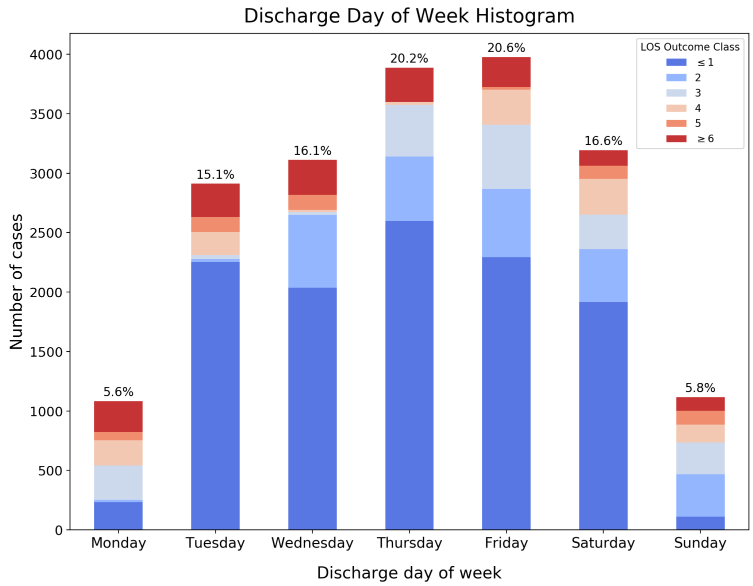


Figure 3.26: Discharge day of week histogram. $x-$axis is the day of week. $y-$axis is the number of surgical cases discharged during each day of week. Each bar is stratified in color by the actual LOS outcome and labeled on top with the percentage of cases within each day of week group.

and Sunday rarely have admissions. The same reasoning applies to the lack of discharges on Wednesday with an LOS of three or four nights, and the scarcity of discharges on Thursday with an LOS of four or five nights.

3.6.4 Readmission Indicator

In addition to the day and time features of admission, we further explore the relation between LOS outcome and 30-day and 60-day readmission respectively. Typically, after discharge, patients would not come back to the hospital as inpatient within a few months at the minimum. However, if their conditions deteriorate or encounter adverse events, they might be re-admitted to the hospital for further surgical operations. In Fig 3.27, we examine the impact of 30-day and 60-day readmission on the LOS outcome. There are 196 30-day readmissions and 461 60-day readmissions. Both are rare scenarios among the total of 19,281 historical cases. Based on their distribution over the outcome classes, both readmission types show a higher percentage of long LOS for 5 nights and more. However, among the middle range of 2 to 4 nights of stay, their proportion is similar to the non-readmission cases. This implies that the readmission flag may be a good indicator of very long or very short LOS, but not among the middle LOS outcomes.

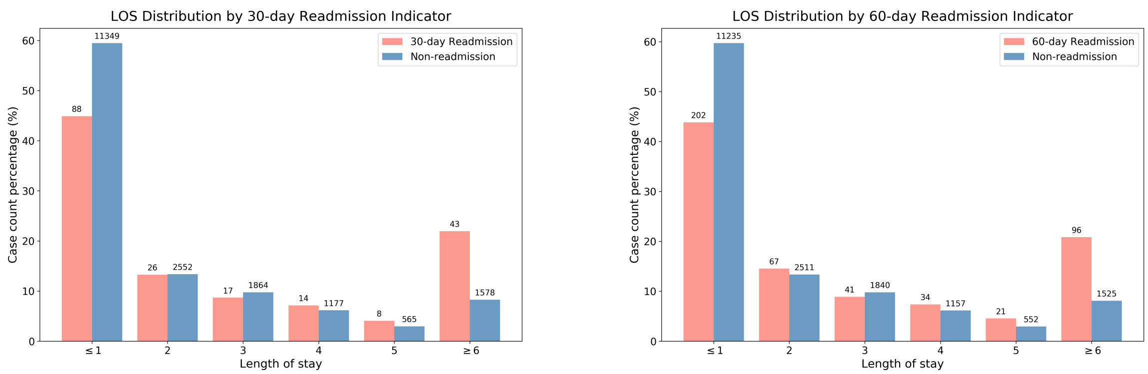


Figure 3.27: 30-day (left) and 60-day (right) readmission and LOS outcome. For both plots, $x-$axis is the LOS outcome class and $y-$axis is the percentage of surgical cases with respect to the full training set under each outcome class. On top of each bar is labeled with the count of surgical cases. Pink represents the cases that have at least one previous discharge within 30 and 60 days respectively. Blue represents the cases without previous discharge in 30 or 60 days.

3.6.5 Operative Length

In an ideal scenario, a smart predictive system would update the remaining LOS dynamically as new information about the patient arrives. In Fig 3.28, we examine the relation between operative length and the post-operative LOS. The box plot shows that shorter operative length implies a short post-operative LOS, while longer operation length is correlated with long LOS. Among the cases with less than 2 nights of post-op stay, over 75% cases underwent an operation that lasted for less than three hours. On the other hand, among the surgical cases with a post-op LOS of at least 3 nights, over 75% of the cases underwent an operation of 3 hours or more. A length of three hours effectively separates the post-op outcome class ≤ 1 from class 3 and above. Therefore, operative length could be a useful feature for post-operative LOS prediction.

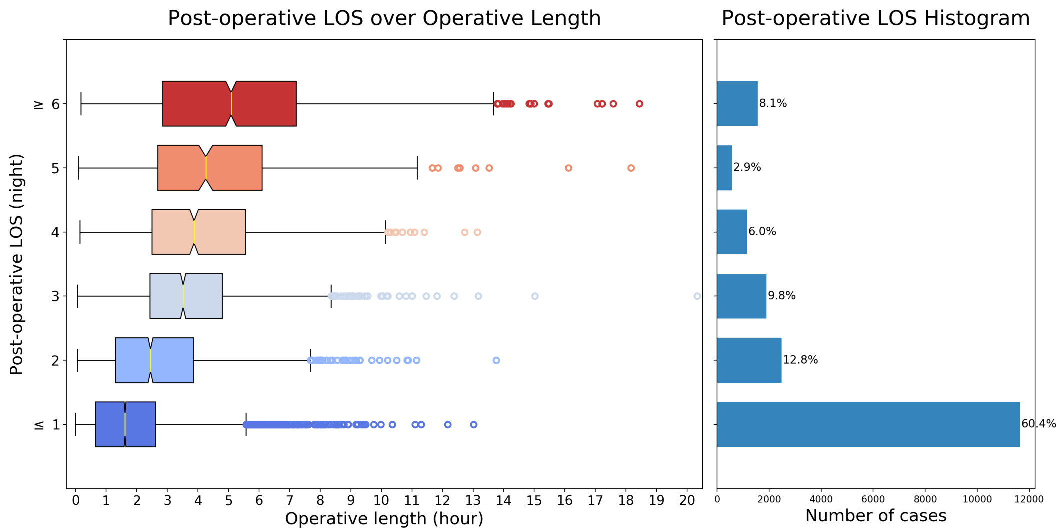


Figure 3.28: Operative length and post-operative LOS. Left is the box plot of surgical length over the post-operative LOS outcome classes. $x-$axis the length of surgical procedure measured in hours. $y-$axis is the post-operative LOS outcome defined as the number of hospitalized nights after surgery. Right plot is the histogram of the post-operative LOS outcome classes. The $x-$axis is the number of surgical cases. Each bar is labeled with the percentage of cases within each outcome class.

3.7 Summary

Based on the exploratory analysis of the features in the training set, we identify a few potentially strong predictors of LOS:

- surgical procedure indicator and the summary statistics related to LOS
- patient care class
- readmission indicator
- length of surgery

Other features, such as major region of residence, chronic conditions and ethnicity, have moderate power to differentiate between very short and very long LOS. None of the features explored so far could untangle the middle range of outcomes on its own.

Chapter 4

Predictive Modeling on LOS

In this chapter, we introduce a machine learning approach that learns a mapping from the patient specifics to one of the LOS outcomes. We start with the problem formulation in Section 4.1 and then describe the end-to-end predictive framework in Section 4.2. We further elaborate on the feature engineering step in Section 4.3. In Section 4.4, we present a brief summary on each machine learning model investigated in this study and in Section 4.5, we describe the ensemble techniques that aggregate the prediction from each individual model.

4.1 Problem Formulation

In this study, we treat LOS as a categorical variable representing the number of nights a patient would spend in the hospital (see definition details in Section 2.3). We aim to predict whether the number of hospitalized nights would be one of the following categories: 0–1, 2, 3, 4, 5, ≥ 6. We formulate this problem as a multiclass classification task where the model is to predict one of the outcome categories based on the input information.

Since the number of nights is an ordinal quantity where there is an inherent order between classes, we also reduce the prediction resolution to investigate five binary classification tasks formulated according to the multi-class outcome categories. In particular, we apply ML models to predict whether the patient would be hospitalized for more than one up to five nights respectively.

4.2 Predictive Framework

Fig 4.1 is the diagram of our predictive framework with three stages: data acquisition, data wrangling and statistical modeling.

In the first stage, we collect the patient data from different sources in the hospital information database according to the inclusion criteria described in Section 2.1. Surgical inpatient visits from January 01, 2018 to September 30, 2021 are used for training statistical models. Cases from October 1, 2021 to March 30, 2022 are treated as an out-of-sample test set to evaluate the model performance. We use 80% of historical data for model training and the rest 20% for validation.

In the second stage, we preprocess the retrieved data into numerical matrices with three steps. We first clean the dataset by removing missing data or replacing them with the median value. During feature engineering, we apply one-hot encoding on categorical features and augment the dataset with new features, such as the summary statistics of clinical variables based on their LOS distribution in the training set. Finally, we denoise the dataset by removing the data points with identical feature values but different LOS outcomes. Optionally, we apply scaler to standardize the processed data matrix. Meanwhile, we save all the meta information, such as input scalers and the observed set of procedures for preprocessing the test and validation set.

In the modeling stage, we train machine learning models on the training data, and evaluate the model performance on validation set preprocessed in the same way as the training set. We perform 5-fold cross validation to select hyperparameters of all models. Then, we repeat data preprocessing on the full historical cases and retrain the selected models, saving the meta information and trained models.

At the inference stage, we use the meta information saved in the previous stage to preprocess the out-of-sample test set. We skip the surgical cases with unseen diagnosis or surgical procedures. Finally, we apply the pre-trained models to predict the LOS of the test cases.

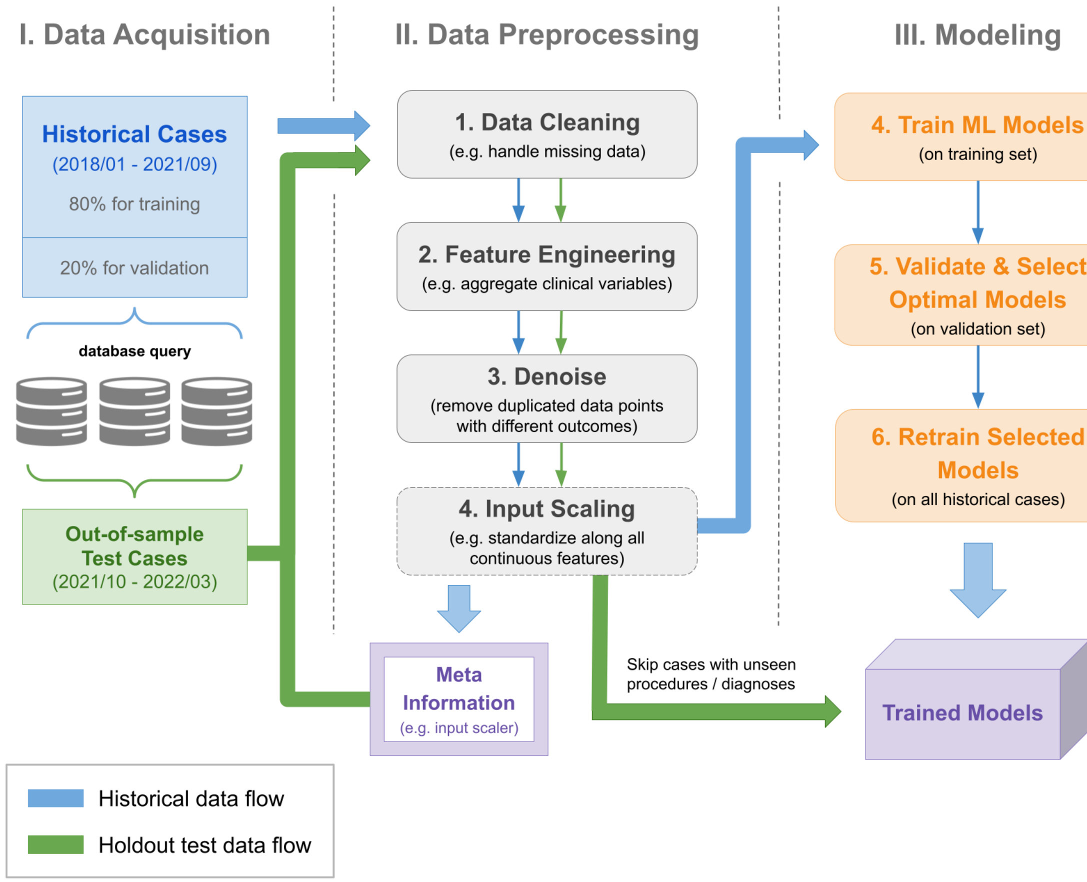


Figure 4.1: LOS prediction framework diagram.

4.3 Data Preprocessing

Data preprocessing consists of four steps: data cleaning, feature engineering, data denoise and input scaling. In this section, we discuss the concepts and details of each step.

4.3.1 Data Cleaning

In this step, we check for missingness across all the features discussed in exploratory data analysis. Our dataset contains missing information in patient's interpreter need, state code and miles traveled. We discard the data points with missing interpreter need and state code, and replace the missing distance with the median miles traveled of the same state or major region. Among 19,281 training samples, we remove 25 cases with missing state code and 154 with missing interpreter need, retaining a total of 19,201 patient encounters.

4.3.2 Feature Engineering

We start by one-hot encoding all categorical variables, such as gender, presence of each CCSR diagnosis code, primary procedure, etc., but there are thousands of categories among the clinical variables. To maximize the predictive power of the feature set while avoiding overwhelming the models with a huge yet sparse input matrix, we manually reduce the high-dimensional binary indicators to a small set of synthesized features.

Specifically, for all clinical variables (procedure type, diagnosis and medication), we synthesize their indication of LOS outcome into the summary statistics of the LOS distribution within each clinical cohort. Fig 4.2 shows the diagram of the feature engineering process. Formally, suppose $X \in R^{n \times d}$ is the data matrix, $y \in R^n$ is the a vector of the continuous LOS and $P = \{p_1, p_2, ..., p_M\}$ is the set of all categories of a clinical variable P (e.g. primary procedure). Let X_{p_i} denote the inpatient cases whose clinical variable is of type p_i and y_{p_i} denote the corresponding LOS values. We compute the following five summary statistics as

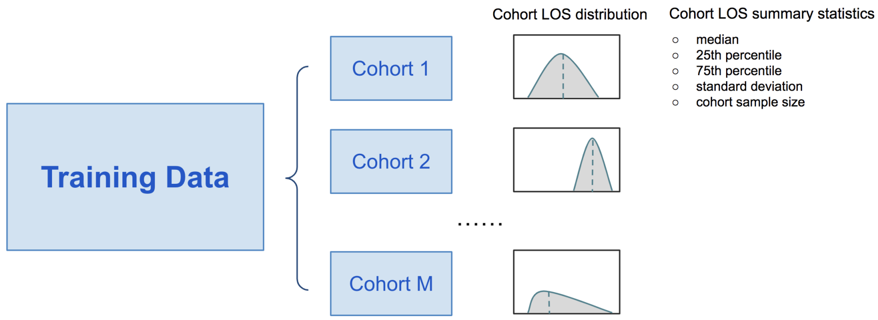


Figure 4.2: Feature engineering of cohort-wise summary statistics.

a synthesized feature vector of p_i:

$$\begin{aligned}
\mathrm{Med}(p_i) &= \mathrm{median}(y_{p_i}) \\
\mathrm{Q25}(p_i) &= \mathrm{quantile}(0.25,\, y_{p_i}) \\
\mathrm{Q75}(p_i) &= \mathrm{quantile}(0.75,\, y_{p_i}) \\
\mathrm{SD}(p_i) &= \sigma(y_{p_i}) \\
\mathrm{Size}(p_i) &= |y_{p_i}|
\end{aligned}$$

In this way, we reduce the thousands of binary indicators for each clinical variable to a vector of five cohort-specific summary statistics. It is important to note that multiple CPT codes, CCSR diagnoses and medication categories could occur in one inpatient visit. In these scenarios, we apply an arbitrary aggregation function along each summary statistics.

4.3.3 Data Denoise

To further improve the data quality, we check for data points with identical feature vector but having diverging outcome values. For example, a few surgical cases share the exact same demographics, procedural and diagnosis characteristics, but fall under different LOS outcome classes. Since machine learning models are designed to learn a function map from the feature space to the outcome, such noise, which violates the definition of mathematical functions, might influence the models in a negative way and should be removed before model training. Among the 19,281 data points in the training set, we remove a total of 36 such cases.

4.3.4 Input Scaling

Features of different scales may pose additional challenges to machine learning models. In this study, we consider three types of input scaling: standardization, min-max scaling and robust scaling.

Standardization

Standardization, also called z-score normalization, scales the value of each feature to have zero-mean and unit variance. This normalization technique is widely used in distance-sensitive algorithms like support vector machines, logistics regression, k-nearest neighbors, etc. Denote $x \in R^n$ as the original data vector of a feature, $\bar{x}$ as the mean of x and σ the standard deviation, standardization transforms x as follow:

$$x' = \frac{x - \bar{x}}{\sigma}$$

Mean and standard deviation of each feature are stored for scaling the test data in prediction.

Min-max Scaling

Min-max scaling is a normalization technique that subtracts the minimum of a feature and scales the feature range to $[0, 1]$:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Minimum and maximum value of each feature are stored for scaling the test data in prediction.

Robust Scaling

Robust scaling normalizes the each feature using statistics that are robust to outliers. It first subtracts the median and then scales the data according to the IQR range:

$$x' = \frac{x - \text{median}(x)}{\text{IQR}(x)}$$

Median and interquartile range of each feature are stored for scaling the test data in prediction.

4.4 Classification Models

In this study, we formulate the LOS prediction as a classification task where models are trained on observed, historical data to infer the outcome category on unseen test data. This section gives an overview of ten classifiers investigated in this study and describes the traits of their hyperparameters.

4.4.1 Logistic Regression (LOGREG)

Logistic regression is a statistical model to predict the probabilities of two or more possible discrete outcomes, given a set of explanatory variables which can be either continuous or categorical. The probability of a data point $X_i \in R^d$ falling under class c is modeled via the softmax function:

$$\Pr(y_i = c) = \frac{e^{\beta_c \cdot X_i}}{\sum_{k=1}^{K} e^{\beta_k \cdot X_i}}$$

where K is the total number of outcome classes and $\beta_k \in R^d$ is a vector that linearly combines the features of a data point. The unknown parameters in β_k are typically jointly estimated by maximum a posteriori (MAP) estimation with regularization of the coefficients in β_k. This method is a well-received baseline model because it provides a probabilistic interpretation of the outcome and does not make assumptions about the distributions of classes in feature space.

In case of a high-dimensional dataset, ℓ-1 or ℓ-2 regularization could be added to prevent overfitting. The regularization strength is an important hyperparameter in model tuning. A stronger regularization leads to sparser coefficient matrix where less important features are assigned a very small weight in prediction.

4.4.2 K-Nearest Neighbors (KNN)

K-nearest neighbor is a machine learning algorithm that predicts the outcome based on the k nearest neighbors of a data point. For a given input instance, it searches through the entire dataset for the closest k data points considered as neighbors for the input. The predicted outcome is the majority class among the k neighboring observations. There are multiple distance metrics to measure the proximity between data points. Euclidean distance and Manhattan distance are commonly used on continuous and discrete values respectively.

The neighborhood scope (value of k) and the distance metrics could be tuned to maximize the performance of this algorithm. Moreover, this algorithm is based on and thus sensitive to distance. When the input features are of different scales, similar data points might be interpreted as far apart when they differ in a particular feature with a wide range. Therefore, normalization techniques could be used to transform the inputs onto a similar scale to ensure performance.

4.4.3 Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised machine learning algorithm that can be used for both regression ans classification. It constructs hyperplane, or a set of hyperplanes, a high- or infinite-dimensional space to separate the data into classes [26]. It optimizes for parameterizing a hyperplane that has the largest margin between classes. Sometimes, when the dataset is not separable by low-dimensional hyperplanes, SVM uses different kernel functions to encode the point-wise distance of a high-dimensional space, where RBF, polynomial and sigmoid kernels are common alternatives to linear kernels. It makes a prediction on a new data point based on which side of the hyperplane the data falls on.

This algorithm is also sensitive to distance, so input normalization could help improve its performance. Additional fine tuning includes adjusting its regularization strength and varying the kernel parameters.

4.4.4 Naive Bayes (NB)

Naive Bayes is a probabilistic classifier based on applying Bayes' theorem. It assumes the features are independent from each other and conditioned on the outcome class. The posterior probability of a data point $x = (x_1, x_2, ..., x_d)$ falling under class k can be decomposed as

$$\Pr(y = k \mid x) = \frac{\Pr(k)\ \Pr(x \mid y = k)}{\Pr(x)}$$

Under the above independence assumption, the posterior probability could be written as the joint probability of all d feature dimensions, and estimated via MAP:

$$\Pr(y = k \mid x) = \frac{1}{Z}\Pr(k)\prod_{i=1}^{d}\Pr(x_i \mid y = k)$$

$$\hat{y} = \operatorname*{argmax}_{k=1,2,...,K} \Pr(y = k \mid x)$$

where Z is a scaling constant. The likelihood of data under a given class, $\Pr(x \mid y = k)$, could assume a Gaussian, multinomial or other families of distributions. When the dataset suffers from class imbalance, Rennie *et al.* proposed the complement Naive Bayes (CNB), a variant of regular multinomial Naive Bayes (MNB) that uses data from all classes except class k to estimate the likelihood under class k [27].

4.4.5 Decision Tree (DT)

Decision tree, also named as Classification and Regression Trees (CART), is defined by recursively partitioning the input space and defining a local model in each resulting region. It uses a flowchart-like structure where each node represents a "test" of an attribute (e.g. whether a a patient is over five years old) and each branch represents the outcome of that test. Each leaf node stands for a class label indicating a final decision along all paths (decision rules) to the leaf. The tree is constructed by recursively selecting a worth-splitting node and partitioning the data according to the criteria. In the classification context, misclassification rate, entropy or Gini index are common error measures for evaluating a proposed partition [28].

A common problem of this model is the tendency to overfit the training data, which is a result of having too many branches due to outliers and irregularities in the data. To remedy this problem, one could stop growing the tree if the decrease in error is not sufficient to justify the extra complexity of adding a branch [28]. To avoid overfitting the training data, decision tree can be pruned by adjusted the values of its maximum depth, minimum number of samples per partitioned region or split, maximum number of features or leaf node, etc.

4.4.6 Bagging Classifier (BAG)

Bagging, also called "boostrap aggregating", is a technique that repeatedly fits base classifiers on random subsamples of data chosen with replacement [29]. It makes a prediction that aggregates the decisions (either by averaging or voting) from the fitted base classifiers. Decision trees often serve as its base learners to generate more stable estimates.

Bagging models can be tuned by proper selection of the maximum subsampl size, the maximum number of features to fit each base learner, the number of times the data is bootstrapped and all the hyperparameters of its base classifiers.

4.4.7 Random Forests (RF)

Random forest is an ensemble model that extends from bagging with decision trees [30]. Bagging models re-run the same algorithm on different subsets of data, so the fitted base predictors can be highly correlated, which limits the amount of variance reduction [28]. Random forest model tackles this issue by fitting a lot of decision trees with random subsets of the dataset while limiting the number of features during each fit. It chooses a random subset of features for each candidate split in the base tree learner. For a classification task with d features, it typically uses $\sqrt{d}$ features in each split.

To optimize the performance of random forest, the following hyperparameters could be tuned: the number of base tree learners, the maximum number of samples to train each base learners and all the hyperparameters in decision tree.

4.4.8 Extremely Randomized Trees (Extra-Trees)

Extremely randomized trees, or Extra-Trees, build upon random forest with one additional step of randomization. It differs from random forest in two ways: first, it uses the full original dataset rather than boostrapped replicas to fit the tree learners; second, in selecting the cut points to split tree nodes, random forest chooses the optimum split based on information gain or Gini impurity, while Extra-Trees chooses a random point. Its hyperparameters can be tuned in the same way as random forest.

4.4.9 Boosting (BST)

Boosting is a popular method of converting a chain of weak learners into a strong model as a whole. The algorithm applies weak learners sequentially to a weighted version of the dataset, where misclassified data points in the earlier iterations are assigned larger weights [28]. The weak learner could be any regression or classification model, but a shallow decision tree considered the best both in terms of accuracy and in producing well-calibrated probabilities

[28, 31].

After each iteration of fitting a base learner, a loss is calculated based on the actual outcome and the predictions by the base learners up to that iteration. Choice of the loss function leads to variants of boosting models. AdaBoost (BST-ADA) uses an exponential loss, LogitBoost uses log loss, gradient boosting machine (GBM) and XGBoost (a more regularized and accelerated version of GBM) uses the absolute error or any other differentiable function as its loss function. Boosting trees can be fine-tuned via adjusting the number of iterations, the regularization strength and the configuration of the base tree learners.

4.5 Ensemble of Strong Classifiers

4.5.1 Voting Classifier

Voting classifier, also named as the majority-rule classifier, is a simple ensemble method that makes a prediction based on the majority decision by its "voter" classifiers. This technique can be useful for a set of equally well performing models in order to balance out their individual weakness [32]. In this study, we use this method to combine all the models discussed above.

4.5.2 Super Learner

Super learner is another ensemble method that makes aggregated prediction based on its base learners. Unlike voting classifier which can directly use the pretrained base classifiers, super learner algorithm first uses cross validation to estimate the performance of the base classifiers. It then learns a weighted combination of the base learners according to their cross validation performance, such that stronger candidates are assigned larger weights [33].

Chapter 5

Performance Evaluation

In this chapter, we present the performance of the classifiers discussed in the previous section. In Section 5.1, we define the evaluation metrics used for both multiclass and binary classifiers. In Section 5.2, we present the prediction performance of 12 off-the-shelf machine learning models and two customized ensembles of several strong candidates. Our best model achieves a test accuracy of 68.8% in LOS multiclass prediction, outperforming surgeon's prediction by over 3%. We further investigate five independent binary classification tasks as an extension of the multiclass problem in Section 5.3. We discover that all models achieve high performance on binary classification based on a low LOS threshold, while the performance drops sharply in the precision-recall tradeoff as the LOS threshold increases.

5.1 Evaluation Metrics

For multiclass classifiers, we use the evaluation metrics summarized in List 5.1 below. Since the outcome class is an ordinal variable, prediction error is not a purely binary indicator of right or wrong. For example, an error off by four days is not identical to that off by one day. Therefore, in addition to the classification accuracy, we propose accuracy with one-day error tolerance ($\epsilon = 1$) as well as mean absolute error to evaluate the classifier's sharpness in prediction.

Accuracy: percentage of correctly classified instances

Accuracy ($\epsilon = 1$): percentage of correctly classified instances with an error tolerance of one night

Mean absolute error (MAE): the average of the absolute prediction errors

Overprediction rate: percentage of instances whose predicted LOS exceeds the actual LOS

Underprediction rate: percentage of instances whose predicted LOS is shorter than the actual LOS

We further visualize the **confusion matrix** where each row represents the instances in an actual class and each column represents the instances in a predicted class. Confusion matrix of a perfect classifier is diagonal.

In correspondence with the six, ordinal outcome classes of the multiclass classification, we examine five binary classification tasks aiming to predict whether a patient would be hospitalized for longer than one, two, ..., up to five days. Their evaluation metrics include recall, precision, receiver operating characteristic (ROC) curve with AUROC and precision-recall curve (PRC) with AUPRC as error metrics defined in List 5.1.

Recall: percentage of true positives among all actually positive instances

Precision: percentage of true positives among all positive predictions

AUROC: area under the receiver operating characteristic, where ROC of a classifier is a curve along the axes of true positive rate and false positive rate as the classifier's discrimination threshold changes

AUPRC: area under precision-recall curve, where PRC of a classifier is a curve along the axes of precision and recall as the classifier's discrimination threshold changes

5.2 Multiclass Classification

Under the multiclass classification setting, we train and evaluate twelve machine learning models and two ensemble models of the strong classifiers. All models are trained on the full historical dataset with a maximum of 785 features. Their hyperparameters and input scalers are selected based on five-fold cross validation. Details of hyperparameter tuning and model selection can be found in the Appendix. For performance evaluation, we use 1471 cases from the out-of-sample test data which have surgeon's prediction. We compare the models' performance with surgeon's prediction under the evaluation metrics described above.

5.2.1 Out-of-sample Test Performance

Table 5.1 summarizes the LOS prediction performance of the multiclass classifiers and surgeon's estimation. Tree-based classifiers like random forest, ExtraTrees and boosting methods have an accuracy of over 67% and a MAE of lower than 0.6. With an error tolerance of one day, these classifiers make correct prediction for over 85.5% cases. All these classifiers tend to underpredict almost twice as often as they overpredict. In comparison, logistic regression and k-nearest neighbors demonstrate slightly worse performance in terms of accuracy and average error, but KNN has a more severe underprediction rate. Support vector machine and naive bayes models have worse performance with a lower accuracy ($< 65\%$), and higher MAE (> 0.66).

Voting ensemble and super learner are customized ensemble models of strong classifiers. Both have comparable performance that is top among rest of the individual classifiers. Voting ensemble achieves the highest accuracy of 68.8%, but it is slightly inferior to super learner in the balancing between over- and under-prediction.

Surgeon's estimation has an accuracy of 65.3%, which is lower than most of the machine learning classifiers. Similar to the classifiers, surgeons tend to underpredict (21%) much more often than they overpredict (13.7%). However, clinicians prediction exhibits the highest accuracy of 88.4% under one-day error tolerance, and has the lowest MAE of 0.54. Such indicates that surgeons are less precise than the models in prediction, but are sharper at identifying the range of patient hospital length of stay, that is, if they make mistakes, they err by a little.

Table 5.1: Out-of-sample Test Performance of Multiclass LOS predictors

Model	**Accuracy**	**Accuracy** ($\epsilon = 1$)	**MAE**	**Overprediction**	**Underprediction**
LGR	66.2%	85.2%	0.61	12.0%	21.8%
KNN	65.9%	85.4%	0.62	9.0%	25.1%
SVM	64.4%	82.7%	0.69	9.9%	25.7%
GaussianNB	60.6%	81.5%	0.73	18.6%	20.9%
MNB	64.7%	82.6%	0.68	13.4%	21.9%
CNB	64.9%	84.1%	0.67	13.6%	21.5%
DT	66.9%	84.4%	0.63	10.4%	22.7%
RF	67.6%	85.9%	0.60	11.6%	20.7%
ExtraTrees	68.1%	85.5%	0.60	11.0%	20.9%
BAG-DT	67.9%	85.9%	0.59	11.5%	20.6%
BST-ADA	67.7%	85.5%	0.61	11.7%	20.6%
XGB	67.8%	86.0%	0.59	12.0%	20.2%
Voting Ensemble	**68.8%**	**85.8%**	**0.58**	**9.9%**	**21.3%**
Super Learner	**68.1%**	**86.1%**	**0.58**	**11.4%**	**20.5%**
Surgeon	65.3%	88.4%	0.54	13.7%	21.0%

5.2.2 Confusion Matrix

To obtain a more transparent picture of surgeon and model's performance, we visualize their respective confusion matrix. Fig 5.1 is the confusion matrix of surgeon's prediction. Each grid is normalized by the total number of cases under the true outcome class (row sum). In other words, the diagonal entries are the recall of each class. The confusion matrix is close to being diagonal, where most of its diagonal entries are darker than their off-diagonal neighbors. The recall rate is the highest in class 1, while falls below 40% among class 2 to ≥ 6. It implies that clinicians could not perfectly differentiate between the cases with medium to long LOS. Moreover, given the lighter color at the diagonal entries in row 2 and 4, clinicians are more likely to predict one day for cases with an actual LOS of 2, and three days for cases whose actual LOS is 4.

In contrast, the ensemble models demonstrate a different pattern. Fig 5.2 shows the confusion matrix of the voting ensemble and super learner. Both matrices are similar: they have a high recall rate for class 1, 3, and ≥ 6, but underperforms over the cases under class 2, 4, and 5. Although they are equally likely to underpredict as the clinicians, the classifiers exhibit a more centralized underprediction pattern: they frequently predict class 1 and 3.

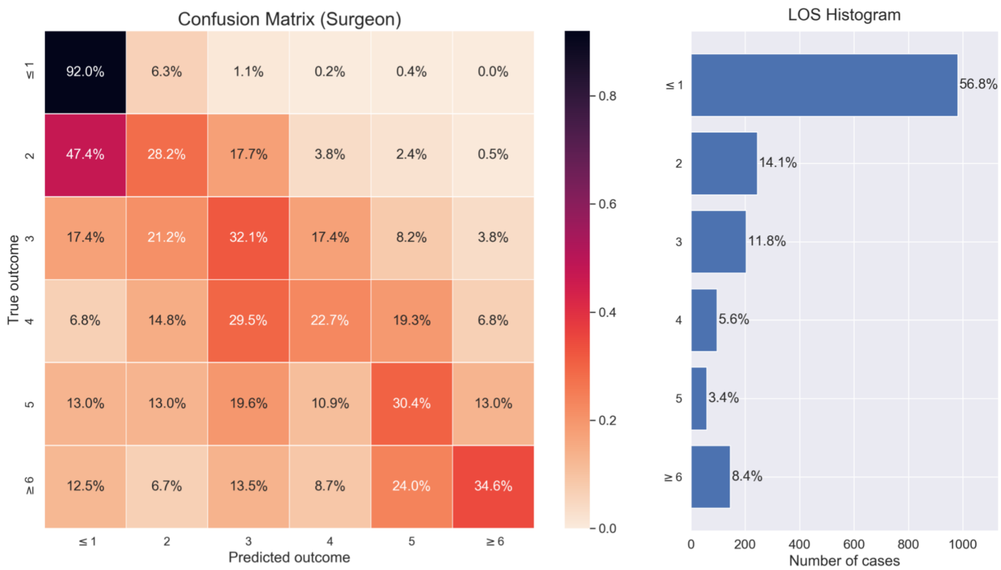


Figure 5.1: Confusion matrix (left) of surgeon's prediction and LOS outcome distribution (right). In the left plot, each row is the true LOS outcome class, and each column is the surgeon-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class.

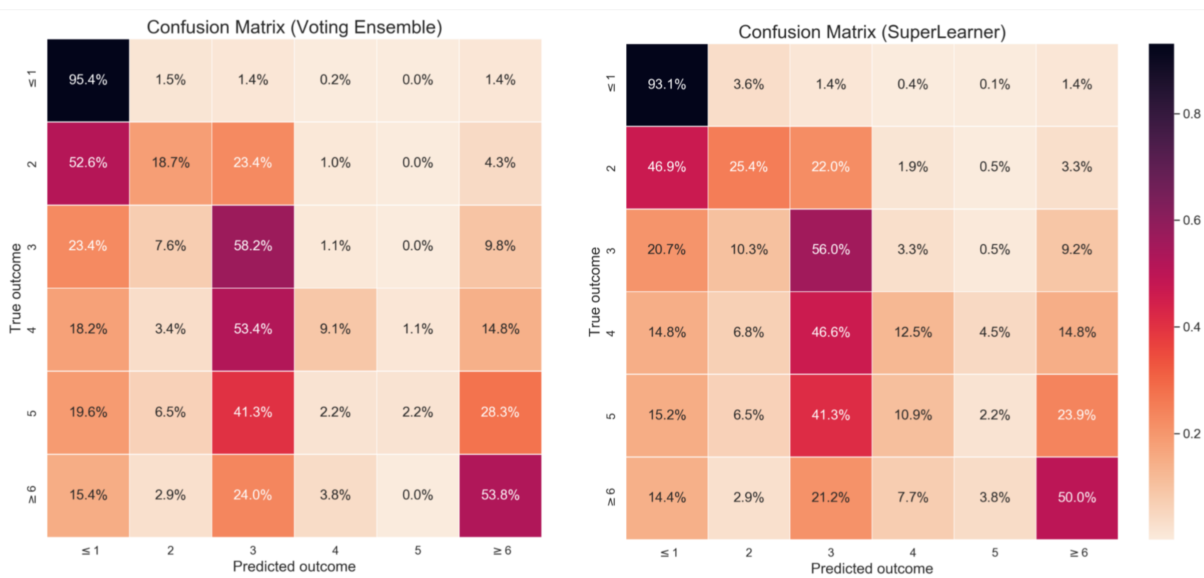


Figure 5.2: Confusion matrix of voting classifier (left) and super learner (right). Each row is the true LOS outcome class, and each column is the model-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted outcome class.

5.2.3 Feature Importance

We leverage Shapley values, which calculates the marginal contribution of each feature to the final prediction, to rank the feature importance [34]. We investigate the feature importance of the voting ensemble model, which is an equally weighted, linear combination of nine base models, by averaging their Shapley values along each feature axis. Fig 5.3 shows the top 15 most important features of the voting ensemble model. Primary procedure summary statistics features are ranked the most important. The corresponding 75th percentile is indicative of all six classes with similar strength, while the 25th percentile is more predictive of very short LOS (class 1) than long LOS (class 2 and above). The LOS standard deviation with each primary procedure cohort is very predictive of long LOS (class 6) but is only moderately important for short LOS (class 3 and below). Summary statistics features by CPT grouping are ranked as the secondary most significant feature set, while miles traveled, diagnosis and medication features are ranked as weaker predictors.

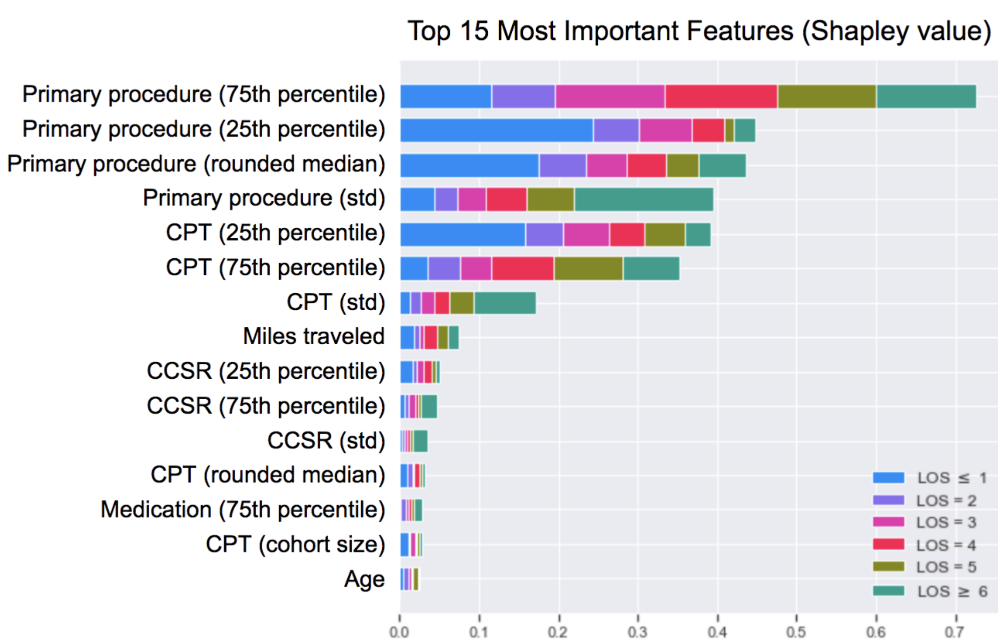


Figure 5.3: Ranking of top 15 most important features. $x-$axis is the Shapley value (larger means more important), and $y-$axis is the feature name. Each bar is the overall Shapley value where the color indicates the contribution of the each feature to predicting the corresponding outcome class.

5.3 Binary Classification

In addition to learning a multiclass classifier, we break down the prediction problem into five independent binary classification tasks. The goal of each is to identify if a patient would stay for strictly more than δ days where $\delta \in \{1, 2, 3, 4, 5\}$. This is a pilot step in investigating the potential of building a hierarchical model where binary classifiers could correct the errors in the multiclass classifiers. Each binary classifier is trained on the full historical dataset with the same features as its multiclass counterpart, and evaluated on the 1471 out-of-sample cases which has a surgeon prediction. Since ROC and PRC relies on the probabilistic estimate of each binary outcome while surgeon's prediction is not a probability, we only compare the models' performance in this section.

5.3.1 AUROC

ROC curves capture the classifier's performance in the trade-off between true positive and false positive rate. A no-skill classifier would predict a random class with equal chance in all scenarios, and thus gives a baseline AUROC of 0.5.

Table 5.2 is a summary of AUROC of the classifiers under each task. Random forest, bagging and XGBoost are the strongest classifiers across all tasks, achieving an AUROC of over 0.88 for all five binary classifications. Both voting ensemble and super learner achieve a similarly optimal performance that is significantly better than a no-skill baseline.

In addition, as the binary threshold increases from 1 to 5, all models show a slight decrease in AUROC. This is likely due to the increase in class imbalance (LOS $> \delta$ becomes more minor) such that it becomes increasingly challenging to precisely infer the true positives while controlling for the false positive rate.

Table 5.2: AUROC of Binary LOS predictors on Test Set

Model	**AUROC** (LOS > 1)	**AUROC** (LOS > 2)	**AUROC** (LOS > 3)	**AUROC** (LOS > 4)	**AUROC** (LOS > 5)
LGR	0.89	0.90	0.88	0.88	0.88
KNN	0.90	0.90	0.87	0.87	0.86
SVM	0.89	0.89	0.87	0.85	0.84
GaussianNB	0.87	0.87	0.84	0.84	0.85
MNB	0.88	0.88	0.86	0.85	0.84
CNB	0.88	0.87	0.82	0.76	0.69
DT	0.87	0.89	0.87	0.86	0.81
RF	0.90	**0.91**	**0.89**	**0.89**	0.89
ExtraTrees	0.90	0.91	0.89	0.89	0.88
BAG-DT	**0.91**	**0.91**	**0.89**	**0.89**	**0.90**
BST-ADA	0.89	**0.91**	0.87	0.88	0.86
XGB	0.89	**0.91**	**0.89**	**0.89**	0.88
Voting Ensemble	0.90	**0.91**	**0.89**	**0.89**	0.88
Super Learner	0.90	**0.91**	**0.89**	**0.89**	0.88

5.3.2 AUPRC

According to the exploratory data analysis on the outcome distribution in 3.1, as the binary cutoff moves from 1 to 5, the size of the minority class (LOS $> \delta$) decreases from 40.7% to 8.4%. In the scenario when correctly classifying the positives is important, PRC with AUPRC is a more appropriate metric as it directly reflects the recall and precision dynamics with the change of the discrimination probability threshold. A no-skill classifier predicts positive class in all circumstances, thus giving an AUPRC equivalent to the ratio of positives.

Table 5.3 shows the AUPRC of the models for each binary classification task. Bagging is the strongest classifier, achieving the highest AUPRCs in all tasks. ExtraTrees and random forest are also strong candidates, reaching a slightly worse performance than bagging classifier.

In predicting whether LOS > 1, almost all classifiers achieve a high AUPRC of over 0.85. As the cutoff increases, there is a significant decay in performance across all classifiers. For example, the AUPRC of bagging classifier drops from 0.91 in the first task to 0.49 in the last prediction task. This means that the classifier is less skilled at inferring whether a surgical case would have a long LOS than a short one, and there is a sharper trade-off between precision and recall. In other words, opting for a high recall leads to a low precision, and vice versa.

Table 5.3: AUPRC of Binary LOS predictors on Test Set

Model	**AUPRC** (LOS > 1)	**AUPRC** (LOS > 2)	**AUPRC** (LOS > 3)	**AUPRC** (LOS > 4)	**AUPRC** (LOS > 5)
LGR	0.87	0.79	0.62	0.51	0.41
KNN	0.89	0.78	0.59	0.48	0.38
SVM	0.87	0.78	0.59	0.48	0.38
GaussianNB	0.84	0.74	0.53	0.43	0.35
MNB	0.88	0.78	0.59	0.46	0.35
CNB	0.88	0.77	0.57	0.39	0.27
DT	0.88	0.80	0.63	0.51	0.40
RF	0.90	0.80	**0.66**	0.56	0.47
ExtraTrees	0.90	0.82	**0.66**	0.56	0.47
BAG-DT	**0.91**	**0.83**	**0.66**	**0.57**	**0.49**
BST-ADA	0.89	0.81	0.61	0.53	0.44
XGB	0.90	0.79	0.60	0.50	0.41
Voting Ensemble	0.90	0.80	0.63	0.52	0.42
Super Learner	0.90	0.79	0.60	0.49	0.41

5.3.3 ROC, PRC

Fig 5.4 shows the ROC and precision-recall curve of all the models for each binary task. All models achieve significantly better performance than a no-skill classifier in all tasks in terms of both ROC and PRC. In the trade-off between true positive and false positive rate, tree-based models and customized ensembles demonstrate a strong performance. In the trade-off between precision and recall, all models show ideal performance in the first task, while exhibit diminishing performance as proportion of the positive class decreases.

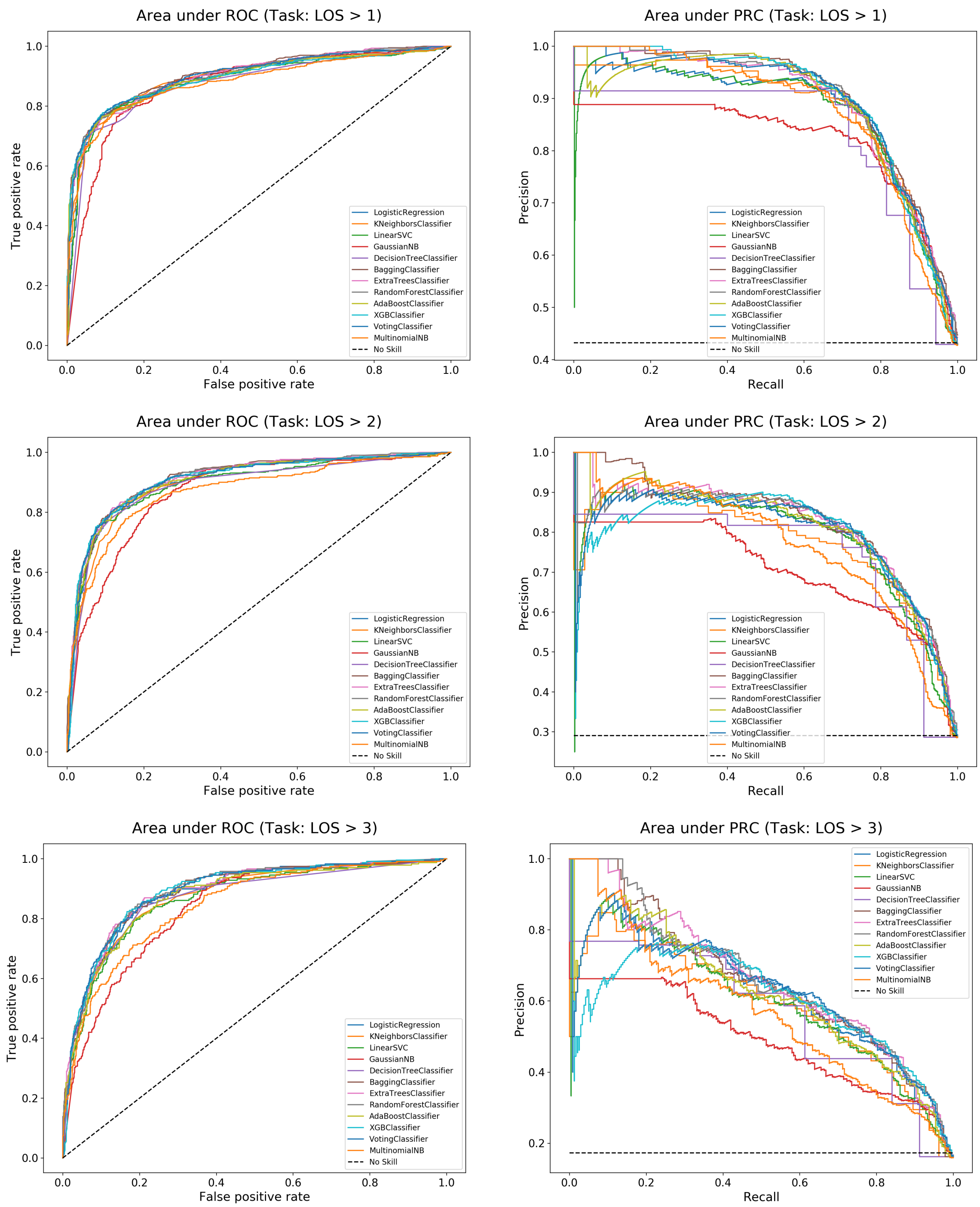


Figure 5.4: ROC curve (left) and PRC curve (right) for five binary classification tasks. Each color represents a classifier specified in the legend. The black dashed line represents the performance of a no-skill classifier.

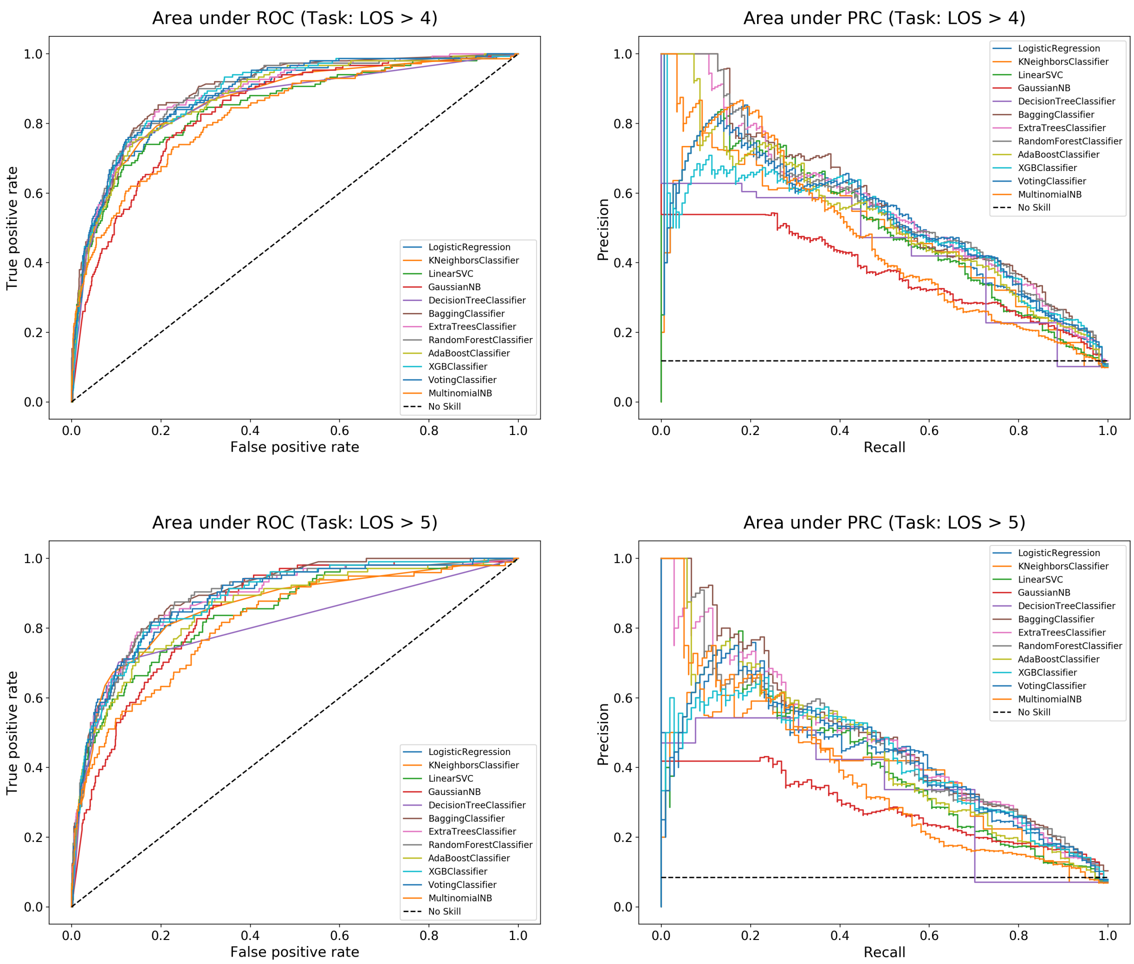


Figure 5.4: ROC curve (left) and PRC curve (right) for five binary classification tasks. Each color represents a classifier specified in the legend. The black dashed line represents the performance of a no-skill classifier.

Chapter 6

Uncertainty Quantification

In the development of physician assistive tools such as an LOS predictor, it is not only imperative to aim for accurate predictions, but also important to quantify the prediction uncertainty, such that the end user could calibrate how much they could trust the tool. In this chapter, we propose three angles to quantify the performance of a prediction. We start with a consensus mechanism that compares the model's estimate with surgeon's prediction in Section 6.1. We discover that when model and surgeon agree, the prediction performance reaches 80.5% in accuracy and 93.2% with a one-day error tolerance. However, when they disagree, the performance of both is sub-optimal. In Section 6.2 and 6.3, we discuss the stratified performance of the models under two clinical categorizations: surgical service line and CPT group. We find that the prediction quality has a large variance among different clinical cohorts. Therefore, the cohort-specific performance is valuable to the end user in judging how much they should weigh the model's suggestion.

6.1 Model & Surgeon

At the time of on-scheduling LOS prediction, surgeons would first input an estimate of how long the patient would be hospitalized, and then receive a model suggestion. To enable better user interaction, we compare the surgeon and models prediction by checking their consensus in three levels: complete agreement, disagreement by one day, and disagreement by more than one day.

6.1.1 Complete Agreement

Table 6.1 shows the model and surgeon's performance over the cases where they completely agree. Voting ensemble has the highest agreement rate of 72% (1059/1471) and achieves an accuracy of 80.5%. It reaches 93.2% in accuracy under one-day error tolerance with a low MAE of 0.32. In comparison, other models, except for naive bayes, agree with the surgeon on over 70% of the cases, and reaches a similarly high performance.

Table 6.1: Classification Performance on Model-surgeon Agreed Cases

Model	**Count (%)**	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**Overprediction**	**Underprediction**	**MAE**
LGR	1035 (70.2%)	80.0%	93.7%	4.9%	15.1%	0.31
KNN	1034 (70.3%)	80.1%	93.4%	3.7%	16.2%	0.32
SVM	1005 (68.3%)	80.8%	93.6%	3.0%	16.2%	0.32
GaussianNB	920 (62.5%)	81.7%	94.1%	4.0%	14.2%	0.28
MNB	981 (66.7%)	80.4%	92.7%	4.1%	15.5%	0.34
CNB	993 (67.5%)	79.7%	93.2%	4.6%	15.7%	0.34
DT	1050 (71.4%)	79.6%	93.2%	4.1%	16.3%	0.33
RF	1046 (71.1%)	79.9%	93.1%	4.6%	15.5%	0.33
ExtraTrees	1051 (71.4%)	80.3%	93.1%	4.6%	15.1%	0.32
BAG-DT	1047 (71.2%)	80.4%	93.4%	4.4%	15.2%	0.32
BST-ADA	1033 (70.2%)	80.2%	93.2%	4.2%	15.7%	0.32
XGB	1048 (71.2%)	79.8%	93.1%	4.7%	15.6%	0.33
Voting Ensemble	**1059 (72.0%)**	**80.5%**	**93.2%**	**4.1%**	**15.4%**	**0.32**
Super Learner	1051 (71.4%)	79.9%	92.9%	4.4%	15.7%	0.33

Fig 6.1 shows the confusion matrix of the voting ensemble model on the surgeon-agreed cases. Over 72% of the cases have a short LOS of no more than one day. The confusion matrix implies that both model and surgeon are prone to underpredict, which most often happens in predicting an LOS of 1 for class 2, and 3 for class 4.

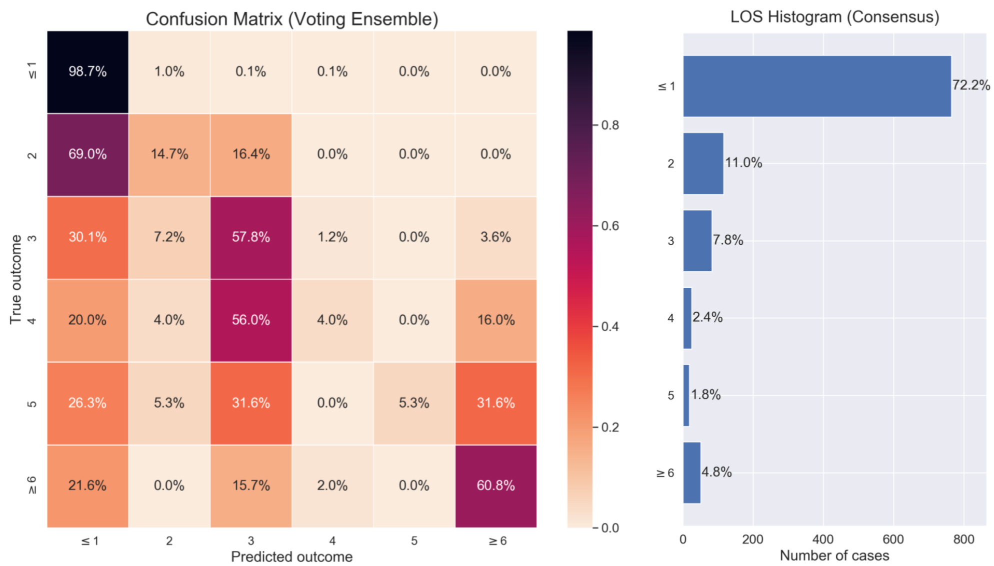


Figure 6.1: Confusion matrix (left) and outcome distribution (right) of cases where Voting Ensemble and surgeon prediction agree. Each row is the true LOS outcome class, and each column is the surgeon-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class.

6.1.2 Disagreement by 1 Day

In the scenario when model and surgeon disagree by exactly one day, the model is more accurate . Table 6.2 lists the performance comparison of surgeon and the fourteen classifiers discussed before. Most of the models disagree with surgeon around 20% of the time, where they significantly surpass surgeon's prediction in accuracy and average error. However, they have moderately worse performance in accuracy with one-day error tolerance.

For instance, voting ensemble disagree with the models by one day among 280 cases in the test set. It achieves an accuracy of 46.4% and MAE of 0.91 among these cases, beating surgeon's accuracy by over 20% and MAE by 0.13. Nevertheless, its accuracy with one-day error tolerance (77.1%) is suboptimal compared with clinician's performance (81.4%). Fig 6.2 shows the confusion matrix of surgeons and the voting ensemble. The model frequently predicts class 1, 3, and 6 and achieves a higher recall in these classes, at the cost of confusing them with class 2, 4 and 5. Surgeons, on the other hand, frequently predict class 2, 4, 5, and rarely predict class 3 and 6 among these cases. Given a relatively even outcome distribution, the high frequency of predicting class 2, 4, 5 potentially buys in more accuracy with one-day error tolerance since these "middle" classes have two neighboring outcomes instead of one for class 1 and 6. Predicting middle ranges more frequently has a higher chance of hitting the true outcome within ± 1 error tolerance.

Table 6.2: Classification Performance when Model and Surgeon Disagree by 1 Day

Model	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**Overprediction**	**Underprediction**	**MAE**	**Count (%)**
LGR	40.4%	75.1%	23.6%	36.0%	0.99	297 (20.2%)
Surgeon	29.0%	81.1%	31.6%	39.4%	1.01	
SVM	38.3%	72.3%	14.0%	47.7%	1.07	264 (17.9%)
Surgeon	29.9%	82.6%	32.6%	37.5%	0.95	
KNN	38.6%	73.7%	20.4%	41.1%	1.01	285 (19.4%)
Surgeon	29.8%	82.5%	34.7%	35.4%	0.98	
GaussianNB	34.5%	77.6%	33.2%	32.3%	0.97	322 (21.9%)
Surgeon	35.4%	82.3%	32.3%	32.3%	0.91	
MNB	39.7%	76.1%	27.6%	32.7%	0.99	297 (20.2%)
Surgeon	32.0%	82.8%	32.7%	35.4%	0.94	
CNB	40.3%	80.0%	29.3%	30.3%	0.92	300 (20.4%)
Surgeon	33.3%	83.0%	32.0%	34.7%	0.91	
DT	44.0%	76.4%	21.5%	34.5%	0.93	275 (18.7%)
Surgeon	25.1%	80.0%	36.7%	38.2%	1.05	
RF	44.7%	79.7%	23.7%	31.7%	0.89	300 (20.4%)
Surgeon	27.3%	81.3%	33.7%	39.0%	1.02	
ExtraTrees	45.1%	78.0%	22.7%	32.2%	0.91	286 (19.4%)
Surgeon	25.9%	81.1%	33.6%	40.6%	1.05	
BAG-DT	44.1%	78.1%	23.6%	32.3%	0.92	297 (20.2%)
Surgeon	26.6%	80.8%	34.0%	39.4%	1.03	
BST-ADA	46.5%	80.1%	23.2%	30.3%	0.87	297 (20.2%)
Surgeon	29.0%	82.8%	35.7%	35.4%	0.96	
XGB	46.1%	80.0%	24.4%	29.5%	0.86	295 (20.1%)
Surgeon	26.4%	82.0%	35.3%	38.3%	1.02	
Voting Ensemble	**46.4%**	**77.1%**	**19.6%**	**33.9%**	**0.91**	**280 (19.0%)**
Surgeon	**24.3%**	**81.4%**	**35.0%**	**40.7%**	**1.04**	
Super Learner	46.2%	79.6%	23.4%	30.4%	0.86	299 (20.3%)
Surgeon	25.8%	82.3%	36.1%	38.1%	1.02	

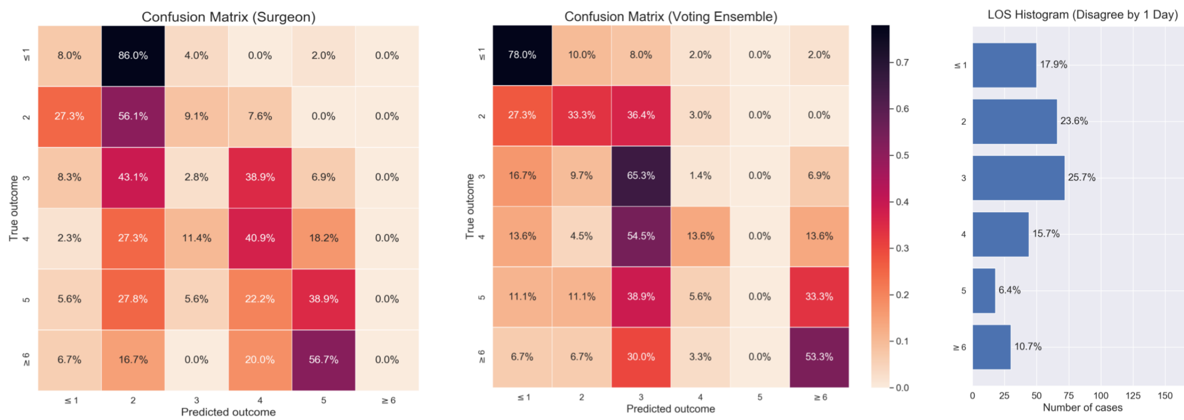


Figure 6.2: Confusion matrix of surgeon prediction (left) and Voting Ensemble (right) when they disagree by 1 day. Each row is the true LOS outcome class, and each column is the surgeon-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class.

6.1.3 Disagreement by > 1 Day

When surgeon and model disagree by more than one day, surgeon's prediction performance significantly exceeds models in both accuracy and sharpness. In Table 6.3, we see that strong disagreement between model and surgeons happens in around 10% of the cases, among which nearly all models have an accuracy of around 20% while surgeons have an accuracy of over 30% when their estimates differ from the respective model prediction by two or more days.

Voting ensemble achieves an accuracy of 22.0% with an average error of 2 days. Among the same set of test cases, surgeons estimate is accurate 30.3% of the time and has an average error of 1.25. Fig 6.3 shows the confusion matrix of voting ensemble model on the surgeon-strongly-disagreed cases and the corresponding outcome distribution. The true LOS distribute evenly among the six classes and does not show severe class imbalance. From the confusion matrix, we see that the overall performance of the voting ensemble is equivalent to the percentage of the majority class in numbers. It most frequently predicts class 3 and 6, which are the more frequent classes. In contrast, surgeon's prediction is sharper with a narrower error range, as its confusion matrix is more diagonal.

In this scenario, both surgeon and model fail to generate very precise estimates of LOS, but surgeon's prediction is of higher quality and should be adopted in the final decision making.

Table 6.3: Classification Performance when Model and Surgeon Disagree by > 1 Day

Model	**Accuracy**	**Accuracy (ϵ = 1)**	**Overprediction**	**Underprediction**	**MAE**	**Count (%)**
LGR	18.7%	43.9%	40.3%	41.0%	2.00	139 (9.4%)
Surgeon	33.8%	64.0%	40.3%	25.9%	1.22	
SVM	14.4%	38.1%	34.2%	51.5%	2.14	202 (13.7%)
Surgeon	34.7%	69.8%	42.1%	23.3%	1.11	
KNN	20.4%	44.1%	24.3%	55.3%	1.91	152 (10.3%)
Surgeon	31.6%	65.1%	42.1%	26.3%	1.24	
GaussianNB	12.2%	36.2%	56.3%	31.4%	2.21	229 (15.6%)
Surgeon	41.5%	73.8%	26.2%	32.3%	1.05	
MNB	16.7%	35.1%	42.9%	40.5%	2.16	168 (11.4%)
Surgeon	36.3%	72.6%	36.9%	26.8%	1.04	
CNB	17.0%	33.3%	41.2%	41.8%	2.33	153 (10.4%)
Surgeon	35.3%	67.3%	37.3%	27.5%	1.15	
DT	18.5%	35.6%	34.9%	46.6%	2.23	146 (10.0%)
Surgeon	38.4%	69.2%	39.0%	22.6%	1.10	
RF	20.0%	40.0%	41.6%	38.4%	2.18	125 (8.5%)
Surgeon	34.4%	65.6%	41.6%	24.0%	1.19	
ExtraTrees	21.6%	41.0%	36.6%	41.8%	2.09	134 (9.1%)
Surgeon	32.1%	66.4%	42.5%	25.4%	1.19	
BAG-DT	20.5%	41.7%	41.7%	37.8%	2.12	127 (8.6%)
Surgeon	31.5%	64.6%	42.5%	26.0%	1.24	
BST-ADA	21.3%	39.7%	42.6%	36.2%	2.16	141 (9.6%)
Surgeon	33.3%	64.5%	36.9%	29.8%	1.25	
XGB	19.5%	41.4%	43.8%	36.7%	2.10	128 (8.7%)
Surgeon	36.7%	64.1%	37.5%	25.8%	1.20	
Voting Ensemble	**22.0**%	**44.7**%	**36.4**%	**41.7**%	**2.01**	**132 (9.0%)**
Surgeon	**30.3**%	**64.4**%	**45.5**%	**24.2**%	**1.25**	
Super Learner	19.8%	43.0%	43.0%	37.2%	2.06	121 (8.2%))
Surgeon	36.4%	64.5%	38.8%	24.8%	1.21	

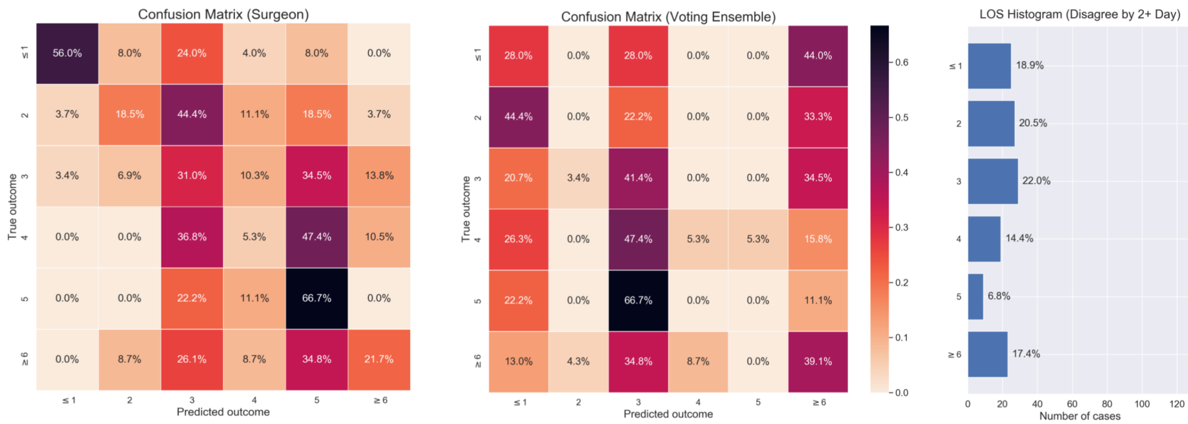


Figure 6.3: Confusion matrix of surgeon prediction (left) and voting ensemble (right) when they disagree by more than 1 day. Each row is the true LOS outcome class, and each column is the surgeon-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class.

6.2 Service Line

Since our patient population contains a wide range of surgical patients who underwent surgical procedures in different specialties, we present a more detailed analysis on the model's performance within each cohort. We first use hospital service line, which is a grouping based on patient encounter attributes such as major disease categories, diagnoses-related groups and ICD codes.

The out-of-sample test set contains 12 service lines. Table 6.4 shows the comparison of model and surgeon's prediction performance within each service line. There is a high variability in the prediction quality across different service lines. For example, both model and surgeon achieve high accuracy in Otolaryngology ($> 87\%$), Plastics ($> 81\%$) and Gynecology ($> 80\%$) cohort. However, they both underperforms in Orthopedics, Neurology and General cohort (accuracy $< 56\%$). This pattern indicates that machine learning model could reach human-like prediction performance when assessed by service line categories. In addition, across all service lines, model has higher accuracy than clinicians, but loses in sharpness since it almost always has a higher MAE and lower accuracy with one-day error tolerance.

Table 6.4: Classification Performance across Surgical Service Lines

Service Line	**Model**	**Count**	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**MAE**	**Overprediction**	**Underprediction**
Orthopedics	Surgeon	389	51.4%	85.1%	0.70	20.3%	28.3%
	Voting Ensemble	389	55.3%	82.3%	0.74	16.5%	28.3%
Otolaryngology	Surgeon	362	87.8%	95.9%	0.20	2.5%	9.7%
	Voting Ensemble	362	89.5%	95.6%	0.19	0.0%	10.5%
General	Surgeon	198	47.0%	77.3%	0.90	26.8%	26.3%
	Voting Ensemble	198	54.5%	73.7%	1.01	23.2%	22.2%
Genitourinary	Surgeon	175	68.6%	89.1%	0.55	6.9%	24.6%
	Voting Ensemble	175	69.7%	85.1%	0.60	4.6%	25.7%
Plastics	Surgeon	162	81.5%	95.1%	0.28	5.6%	13.0%
	Voting Ensemble	162	82.1%	93.8%	0.32	1.2%	16.7%
Neurology	Surgeon	124	42.7%	83.1%	0.82	27.4%	29.8%
	Voting Ensemble	124	50.0%	76.6%	0.92	21.0%	29.0%
Gastrointestinal	Surgeon	27	70.4%	88.9%	0.63	11.1%	18.5%
	Voting Ensemble	27	77.8%	85.2%	0.74	0.0%	22.2%
Gynecology	Surgeon	20	80.0%	100.0%	0.20	10.0%	10.0%
	Voting Ensemble	20	85.0%	100.0%	0.15	0.0%	15.0%
Pulmonary	Surgeon	4	75.0%	100.0%	0.25	0.0%	25.0%
	Voting Ensemble	4	75.0%	75.0%	1.25	0.0%	25.0%
Pain	Surgeon	4	75.0%	75.0%	0.50	0.0%	25.0%
	Voting Ensemble	4	75.0%	75.0%	0.75	0.0%	25.0%
Eye	Surgeon	3	66.7%	66.7%	0.67	0.0%	33.3%
	Voting Ensemble	3	66.7%	66.7%	0.67	0.0%	33.3%
Dental	Surgeon	3	66.7%	100.0%	0.33	0.0%	33.3%
	Voting Ensemble	3	66.7%	100.0%	0.33	0.0%	33.3%

6.3 CPT Group

Hospital service line provides a coarse grouping of the surgical patients based on their diagnostics. To gain a more detailed picture with respect to the surgical procedures, we further present the same analysis on level-3 CPT group cohorts. The test set contains 87 level-3 or below CPT groups, which cover a wide range of procedures based on the organ or body area.

Table 6.5 summarizes the model and surgeon's prediction performance for each of the 87 CPT group cohorts sorted in descending order by their sample size. Model and surgeon achieve comparable performance across all cohorts. The model's prediction is more accurate, but has a wider error range than surgeon's estimate. Among the 87 cohorts, the model has a higher accuracy among 34.5% (30/87) cohorts, it achieve equal accuracy among 42.5% (37/87) cohorts and it has a lower accuracy than clinicians among 17.2% (15/87) cohorts. In terms of average error, the model outperforms surgeons prediction among 40.2% (35/87) cohorts, achieves equivalent performance among 29.9% (26/87) cohorts and underperforms among 24.1% (21/87) cohorts.

Both model and surgeon estimates show varying quality between cohorts. For example, the predictions are more accurate ($> 80\%$) on patients with procedures on tonsils and larynx, whereas they have low accuracy ($< 50\%$) on patients who underwent procedures on the spine and hip joint. In stratifying the patient population by their procedural group, we are able to deliver more transparent evaluation to the end user, such that they could interpret the machine learning suggestions with varying degrees of trust, depending on the cohort-specific accuracy.

Table 6.5: Classification Performance across Level-3 CPT Group

Level-3 CPT Group	**Model**	**Count**	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**Overprediction**	**Underprediction**	**MAE**
Surgical Procedures on the Pharynx, Adenoids, and Tonsils	Surgeon	216	90.3%	94.9%	0.0%	9.7%	0.18
	Voting Ensemble	216	90.3%	94.9%	0.0%	9.7%	0.18
Surgical Procedures on the Trachea and Bronchi	Surgeon	143	79.0%	93.0%	7.0%	14.0%	0.33
	Voting Ensemble	143	82.5%	90.2%	1.4%	16.1%	0.36
Surgical Procedures on the Larynx	Surgeon	135	78.5%	92.6%	6.7%	14.8%	0.32
	Voting Ensemble	135	84.4%	91.9%	0.0%	15.6%	0.27
Surgical Procedures on the Spine (Vertebral Column) Cervical, thoracic, and lumbar spine.	Surgeon	125	40.0%	76.8%	40.0%	20.0%	0.91
	Voting Ensemble	125	47.2%	76.0%	20.0%	32.8%	0.96
Surgical Procedures on the Pelvis and Hip Joint	Surgeon	118	35.6%	85.6%	19.5%	44.9%	0.89
	Voting Ensemble	118	42.4%	80.5%	28.8%	28.8%	0.86
General Surgical Procedures on the Musculoskeletal System	Surgeon	96	46.9%	79.2%	30.2%	22.9%	0.82
	Voting Ensemble	96	53.1%	81.3%	15.6%	31.3%	0.83
Surgical Procedures on the Skull, Meninges, and Brain	Surgeon	84	40.5%	81.0%	31.0%	28.6%	0.86
	Voting Ensemble	84	53.6%	77.4%	15.5%	31.0%	0.83
Surgical Procedures on the Bladder	Surgeon	83	66.3%	86.7%	8.4%	25.3%	0.65
	Voting Ensemble	83	67.5%	81.9%	8.4%	24.1%	0.75
Surgical Procedures on the Middle Ear	Surgeon	79	82.3%	92.4%	1.3%	16.5%	0.34
	Voting Ensemble	79	83.5%	92.4%	0.0%	16.5%	0.33
Surgical Procedures on the Ureter	Surgeon	72	63.9%	84.7%	8.3%	27.8%	0.74
	Voting Ensemble	72	68.1%	80.6%	5.6%	26.4%	0.74
Surgical Procedures on the Breast	Surgeon	65	95.4%	100.0%	0.0%	4.6%	0.05
	Voting Ensemble	65	95.4%	100.0%	0.0%	4.6%	0.05
Surgical Procedures on the Femur (Thigh Region) and Knee Joint	Surgeon	57	64.9%	91.2%	7.0%	28.1%	0.53
	Voting Ensemble	57	64.9%	86.0%	3.5%	31.6%	0.63
Surgical Procedures on the Kidney	Surgeon	54	64.8%	85.2%	9.3%	25.9%	0.57
	Voting Ensemble	54	66.7%	79.6%	3.7%	29.6%	0.63
Surgical Procedures on the Spine and Spinal Cord	Surgeon	49	42.9%	81.6%	26.5%	30.6%	0.84
	Voting Ensemble	49	51.0%	79.6%	22.4%	26.5%	0.82
Surgical Procedures on the Leg (Tibia and Fibula) and Ankle Joint	Surgeon	46	52.2%	91.3%	17.4%	30.4%	0.63
	Voting Ensemble	46	52.2%	78.3%	13.0%	34.8%	0.78
Surgical Procedures on the Head	Surgeon	35	68.6%	91.4%	17.1%	14.3%	0.43
	Voting Ensemble	35	68.6%	91.4%	0.0%	31.4%	0.51
Surgical Procedures on the Stomach	Surgeon	34	35.3%	82.4%	0.0%	64.7%	1.00
	Voting Ensemble	34	67.6%	91.2%	8.8%	23.5%	0.50
Surgical Procedures on the Penis	Surgeon	34	79.4%	97.1%	0.0%	20.6%	0.24
	Voting Ensemble	34	76.5%	97.1%	2.9%	20.6%	0.26
Endoscopy/Arthroscopy Procedures on the Musculoskeletal System	Surgeon	34	64.7%	88.2%	11.8%	23.5%	0.62
	Voting Ensemble	34	67.6%	82.4%	11.8%	20.6%	0.71
Surgical Procedures on the Nose	Surgeon	32	81.3%	96.9%	0.0%	18.8%	0.22
	Voting Ensemble	32	81.3%	90.6%	0.0%	18.8%	0.34
Surgical Procedures on the Extracranial Nerves, Peripheral Nerves, and Autonomic Nervous System	Surgeon	32	46.9%	93.8%	31.3%	21.9%	0.59
	Voting Ensemble	32	56.3%	84.4%	15.6%	28.1%	0.63
Special Otorhinolaryngologic Services and Procedures	Surgeon	32	81.3%	90.6%	3.1%	15.6%	0.44
	Voting Ensemble	32	84.4%	90.6%	0.0%	15.6%	0.41
Surgical Procedures on the Intestines (Except Rectum)	Surgeon	31	32.3%	64.5%	38.7%	29.0%	1.19
	Voting Ensemble	31	32.3%	45.2%	41.9%	25.8%	2.03
Surgical Procedures on the Palate and Uvula	Surgeon	30	83.3%	93.3%	0.0%	16.7%	0.30
	Voting Ensemble	30	83.3%	90.0%	0.0%	16.7%	0.33
Surgical Procedures on the Foot and Toes	Surgeon	30	56.7%	86.7%	13.3%	30.0%	0.63
	Voting Ensemble	30	60.0%	73.3%	3.3%	36.7%	0.83
Surgical Procedures on the Lungs and Pleura	Surgeon	28	32.1%	57.1%	57.1%	10.7%	1.18
	Voting Ensemble	28	53.6%	75.0%	21.4%	25.0%	0.86
Surgical Procedures on the Abdomen, Peritoneum, and Omentum	Surgeon	28	71.4%	85.7%	25.0%	3.6%	0.46
	Voting Ensemble	28	75.0%	89.3%	17.9%	7.1%	0.39
Surgical Procedures on the Rectum	Surgeon	24	20.8%	66.7%	45.8%	33.3%	1.46
	Voting Ensemble	24	33.3%	45.8%	33.3%	33.3%	2.08
Surgical Procedures on the Thyroid Gland	Surgeon	22	81.8%	90.9%	0.0%	18.2%	0.32
	Voting Ensemble	22	77.3%	90.9%	4.5%	18.2%	0.36
Surgical Repair (Closure) Procedures on the Integumentary System	Surgeon	21	66.7%	95.2%	4.8%	28.6%	0.43
	Voting Ensemble	21	71.4%	90.5%	0.0%	28.6%	0.57

Table 6.5: Classification Performance across Level-3 CPT Group

Level-3 CPT Group	Model	Count	Accuracy	Accuracy ($\epsilon = 1$)	Overprediction	Underprediction	MAE
Surgical Procedures on the Esophagus	Surgeon	17	76.5%	88.2%	5.9%	17.6%	0.65
	Voting Ensemble	17	70.6%	76.5%	0.0%	29.4%	1.06
Surgical Procedures on the Inner Ear	Surgeon	15	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	15	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Biliary Tract	Surgeon	15	80.0%	93.3%	6.7%	13.3%	0.47
	Voting Ensemble	15	80.0%	93.3%	6.7%	13.3%	0.47
Surgical Procedures on the Neck (Soft Tissues) and Thorax	Surgeon	14	35.7%	92.9%	42.9%	21.4%	0.79
	Voting Ensemble	14	50.0%	92.9%	7.1%	42.9%	0.57
Surgical Procedures on the Ovary	Surgeon	13	61.5%	92.3%	23.1%	15.4%	0.46
	Voting Ensemble	13	69.2%	92.3%	7.7%	23.1%	0.54
Surgical Procedures on the Lymph Nodes and Lymphatic Channels	Surgeon	13	30.8%	84.6%	38.5%	30.8%	0.92
	Voting Ensemble	13	23.1%	92.3%	30.8%	46.2%	0.85
Surgical Procedures on the Urethra	Surgeon	11	72.7%	90.9%	9.1%	18.2%	0.45
	Voting Ensemble	11	63.6%	81.8%	27.3%	9.1%	0.55
Surgical Procedures on the Shoulder	Surgeon	11	72.7%	90.9%	0.0%	27.3%	0.36
	Voting Ensemble	11	54.5%	72.7%	18.2%	27.3%	1.00
Surgical Procedures on the Lips	Surgeon	11	72.7%	100.0%	9.1%	18.2%	0.27
	Voting Ensemble	11	72.7%	100.0%	0.0%	27.3%	0.27
Surgical Procedures on the Hand and Fingers	Surgeon	11	90.9%	100.0%	9.1%	0.0%	0.09
	Voting Ensemble	11	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Liver	Surgeon	10	90.0%	100.0%	0.0%	10.0%	0.10
	Voting Ensemble	10	70.0%	100.0%	30.0%	0.0%	0.30
Surgery Others	Surgeon	10	50.0%	90.0%	20.0%	30.0%	0.70
	Voting Ensemble	10	60.0%	60.0%	30.0%	10.0%	1.10
Surgical Procedures on the Vagina	Surgeon	9	33.3%	100.0%	0.0%	66.7%	0.67
	Voting Ensemble	9	33.3%	88.9%	22.2%	44.4%	0.89
Surgical Procedures on the Testis	Surgeon	9	77.8%	88.9%	0.0%	22.2%	0.56
	Voting Ensemble	9	77.8%	100.0%	11.1%	11.1%	0.22
Surgical Procedures on the Forearm and Wrist	Surgeon	9	77.8%	88.9%	0.0%	22.2%	0.67
	Voting Ensemble	9	88.9%	88.9%	0.0%	11.1%	0.56
Surgical Procedures on the Corpus Uteri	Surgeon	9	77.8%	100.0%	11.1%	11.1%	0.22
	Voting Ensemble	9	77.8%	88.9%	11.1%	11.1%	0.33
Surgical Procedures on Arteries and Veins	Surgeon	9	55.6%	88.9%	11.1%	33.3%	0.67
	Voting Ensemble	9	77.8%	100.0%	0.0%	22.2%	0.22
Ultrasonic Guidance Procedures	Surgeon	8	87.5%	87.5%	0.0%	12.5%	0.25
	Voting Ensemble	8	87.5%	87.5%	0.0%	12.5%	0.25
Surgical Procedures on the Parathyroid, Thymus, Adrenal Glands, Pancreas, and Carotid Body	Surgeon	8	62.5%	62.5%	37.5%	0.0%	1.00
	Voting Ensemble	8	62.5%	62.5%	12.5%	25.0%	0.88
Surgical Procedures on the Humerus (Upper Arm) and Elbow	Surgeon	8	87.5%	87.5%	0.0%	12.5%	0.25
	Voting Ensemble	8	75.0%	87.5%	12.5%	12.5%	0.38
Surgical Procedures on the Dentoalveolar Structures	Surgeon	7	57.1%	100.0%	0.0%	42.9%	0.43
	Voting Ensemble	7	57.1%	100.0%	0.0%	42.9%	0.43
Surgical Procedures on the Anus	Surgeon	7	28.6%	85.7%	42.9%	28.6%	1.14
	Voting Ensemble	7	0.0%	14.3%	71.4%	28.6%	3.14
Surgical Procedures on the Salivary Gland and Ducts	Surgeon	6	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	6	83.3%	100.0%	0.0%	16.7%	0.17
Surgical Procedures on the External Ear	Surgeon	6	83.3%	100.0%	0.0%	16.7%	0.17
	Voting Ensemble	6	83.3%	100.0%	0.0%	16.7%	0.17
Surgical Procedures on the Accessory Sinuses	Surgeon	6	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	6	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Tongue and Floor of Mouth	Surgeon	5	60.0%	100.0%	0.0%	40.0%	0.40
	Voting Ensemble	5	60.0%	100.0%	0.0%	40.0%	0.40
Surgical Procedures on the Oviduct/Ovary	Surgeon	5	80.0%	100.0%	0.0%	20.0%	0.20
	Voting Ensemble	5	60.0%	100.0%	0.0%	40.0%	0.40
Other Diagnostic Radiology (Diagnostic Imaging) Related Procedures	Surgeon	5	60.0%	80.0%	0.0%	40.0%	0.60
	Voting Ensemble	5	20.0%	40.0%	60.0%	20.0%	2.40
Gastroenterology Procedures	Surgeon	5	80.0%	100.0%	20.0%	0.0%	0.20
	Voting Ensemble	5	80.0%	80.0%	0.0%	20.0%	1.00

Table 6.5: Classification Performance across Level-3 CPT Group

Level-3 CPT Group	Model	Count	Accuracy	Accuracy ($\epsilon = 1$)	Overprediction	Underprediction	MAE
Surgical Procedures on the Mediastinum	Surgeon	4	25.0%	50.0%	50.0%	25.0%	2.00
	Voting Ensemble	4	50.0%	75.0%	25.0%	25.0%	0.75
Other Diagnostic Ultrasound Procedures	Surgeon	4	50.0%	100.0%	25.0%	25.0%	0.50
	Voting Ensemble	4	25.0%	25.0%	75.0%	0.0%	2.25
Fluoroscopic Guidance	Surgeon	4	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	4	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Vestibule of Mouth	Surgeon	3	33.3%	100.0%	0.0%	66.7%	0.67
	Voting Ensemble	3	33.3%	100.0%	0.0%	66.7%	0.67
Surgical Procedures on the Appendix	Surgeon	3	66.7%	66.7%	33.3%	0.0%	1.00
	Voting Ensemble	3	100.0%	100.0%	0.0%	0.0%	0.00
Diagnostic Radiology (Diagnostic Imaging) Procedures of the Lower Extremities	Surgeon	3	66.7%	100.0%	0.0%	33.3%	0.33
	Voting Ensemble	3	66.7%	100.0%	0.0%	33.3%	0.33
Chemotherapy Administration and Other Highly Complex Drug or Highly Complex Biologic Agent Administration	Surgeon	3	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	3	100.0%	100.0%	0.0%	0.0%	0.00
Anesthesia for Procedures on the Head	Surgeon	3	66.7%	100.0%	0.0%	33.3%	0.33
	Voting Ensemble	3	66.7%	100.0%	0.0%	33.3%	0.33
Surgical Procedures on the Spleen	Surgeon	2	50.0%	100.0%	50.0%	0.0%	0.50
	Voting Ensemble	2	50.0%	100.0%	50.0%	0.0%	0.50
Surgical Procedures on the Pancreas	Surgeon	2	50.0%	100.0%	0.0%	50.0%	0.50
	Voting Ensemble	2	0.0%	100.0%	100.0%	0.0%	1.00
Surgical Procedures on the Eyeball	Surgeon	2	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	2	100.0%	100.0%	0.0%	0.0%	0.00
Introduction or Removal Procedures on the Integumentary System	Surgeon	2	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	2	100.0%	100.0%	0.0%	0.0%	0.00
Diagnostic Radiology (Diagnostic Imaging) Procedures of the Urinary Tract	Surgeon	2	50.0%	50.0%	0.0%	50.0%	1.50
	Voting Ensemble	2	100.0%	100.0%	0.0%	0.0%	0.00
Application of Casts and Strapping	Surgeon	2	50.0%	50.0%	0.0%	50.0%	1.00
	Voting Ensemble	2	50.0%	50.0%	0.0%	50.0%	1.00
Surgical Procedures on the Vulva, Perineum and Introitus	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Spermatic Cord	Surgeon	1	0.0%	0.0%	100.0%	0.0%	3.00
	Voting Ensemble	1	0.0%	100.0%	100.0%	0.0%	1.00
Surgical Procedures on the Skin, Subcutaneous and Accessory Structures	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Seminal Vesicles	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Scrotum	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00
Surgical Procedures on the Prostate	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00
General Surgical Procedures on the Hemic and Lymphatic Systems	Surgeon	1	100.0%	100.0%	0.0%	0.0%	0.00
	Voting Ensemble	1	0.0%	100.0%	100.0%	0.0%	1.00
Audiologic Function Tests	Surgeon	1	0.0%	0.0%	0.0%	100.0%	4.00
	Voting Ensemble	1	0.0%	0.0%	0.0%	100.0%	4.00
Active Wound Care Management	Surgeon	1	0.0%	100.0%	0.0%	100.0%	1.00
	Voting Ensemble	1	100.0%	100.0%	0.0%	0.0%	0.00

Chapter 7

Post-admission LOS Prediction

To refine the predictive performance as new patient information becomes available along the patient visit timeline, we apply the multiclass LOS prediction approach at admission time and post-operation. In Section 7.1 and 7.2, we incorporate new features acquired upon admission (admission date time, readmission flag, patient care status) and post-operation (operative length, operation end time) respectively. In the model evaluation, we discover that the post-admission features improve the model performance marginally (+0.2% in accuracy) for on-admission prediction while to a greater extent for post-operative prediction (+1.3% in accuracy). Patient care class, admission date time and operative length features are ranked among the top 15 most important features by Shapley value.

7.1 On Admission

7.1.1 Additional Features

To predict LOS at the time of admission, we engineer a list of features in addition to the feature set for pre-admission prediction:

- admission hour of day
- admission day of week

- 30-day readmission indicator
- 60-day readmission indicator
- patient care class (inpatient vs observation)

7.1.2 Prediction Performance

We apply nine machine learning models (logistic regression, support vector machine, k-nearest neighbors, decision tree, random forest, extra trees, bagging classifier, Adaboost and XGBoost) together with a voting ensemble of them. Table 7.1 shows the model performance. Voting ensemble achieves the best performance in an accuracy (69.0%) while random forest, extra trees and bagging reach a comparable accuracy level with higher accuracy with one-day error tolerance and lower average error. In comparison, surgeon's prediction outperforms the models in sharpness but underperforms in accuracy.

Table 7.1: On-admission LOS Prediction Performance

Model	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**MAE**	**Overprediction**	**Underprediction**
LGR	66.6%	86.7%	0.58	12.5%	20.9%
SVM	66.4%	85.5%	0.61	10.8%	22.8%
KNN	66.3%	85.1%	0.60	8.7%	24.9%
DT	67.6%	85.5%	0.60	11.4%	20.9%
RF	68.9%	87.1%	0.55	11.1%	20.0%
ExtraTrees	68.7%	86.6%	0.55	10.7%	20.6%
BAG-DT	68.5%	87.2%	0.55	10.5%	20.9%
BST-ADA	67.2%	86.0%	0.58	12.4%	20.3%
XGB	68.3%	87.0%	0.55	11.9%	19.8%
Voting Ensemble	**69.0%**	**86.7%**	**0.56**	**10.1%**	**20.9%**
Surgeon	65.3%	88.4%	0.54	13.7%	21.0%

Confusion Matrix

Compared to the pre-admission model performance (Table 5.1), the voting ensemble model is improved in sharpness given its higher accuracy with one-day error tolerance (+0.9%) and lower MAE (-0.02). Fig 7.1 shows the confusion matrix of pre- and on-admission prediction by the voting ensemble model. Even though both demonstrate a strong tendency to underpredict, on-admission model is less likely to confuse very short LOS (class 1) with long LOS (class 3 and above). Moreover, the recall at class 2, 4, 5 is improved by over 2%, which also implies the on-admission model is sharper.

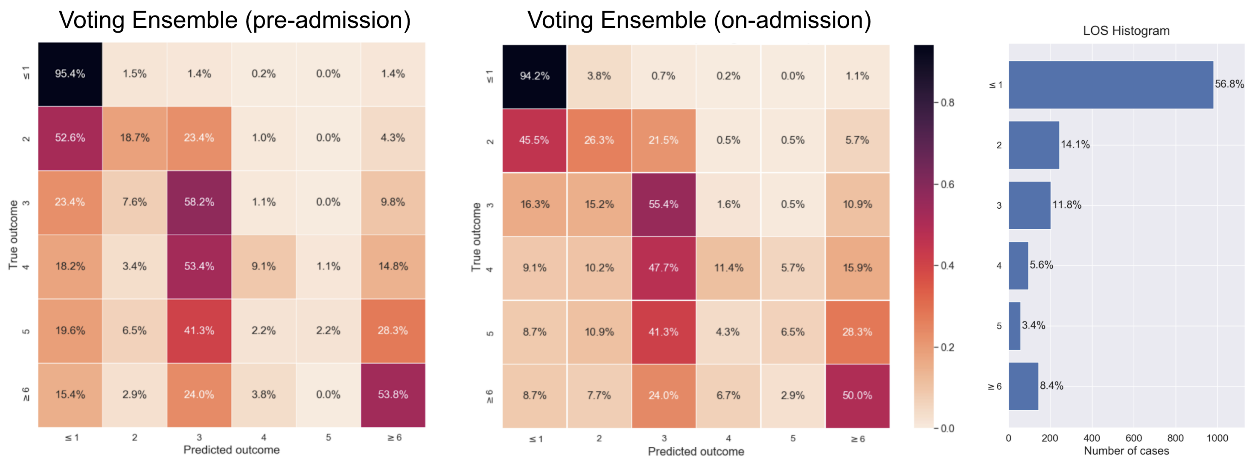


Figure 7.1: Confusion matrix of pre-admission (left) and on-admission (middle) prediction by the voting ensemble model and LOS distribution (right). In the left and middle plot, each row is the true LOS outcome class, and each column is the model-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class. The right plot is the histogram of LOS distribution over the six outcome classes.

Feature Importance

To better understand how the model interprets the features, we use Shapley value to rank features based on their marginal contribution to the model prediction. Fig 7.2 shows the top 15 most important features by the voting ensemble model. Patient care class is considered the strongest differentiator of short LOS (class 2 and below) and long LOS, which aligns with our EDA in Fig 3.22 where almost all observation-class patients were discharged within 2 days, while inpatients-class patients were equally likely to stay for class 3 and above. Procedure-relevant features are the important factors next to care class. The admission date time features are ranked as moderately important. Admission day of week is indicative of very short LOS (class 2 and below) and very long LOS (class 5 and above), while admission hour of day is less indicative of class 2 and 5.

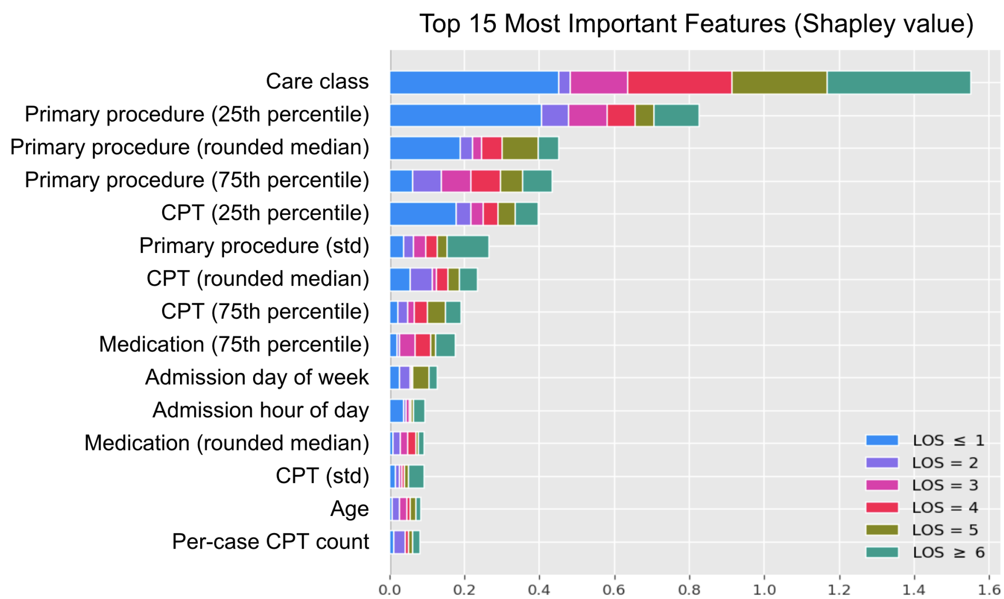


Figure 7.2: Top 15 most important features by voting ensemble model in on-admission prediction. $x-$axis is the Shapley value (larger means more important), and $y-$axis is the feature name. Each bar is the overall Shapley value where the color indicates the contribution of the each feature to predicting the corresponding outcome class.

7.2 Post-operation

7.2.1 Additional Features

For post-operative LOS prediction, we further leverage the observed operation information on each patient encounter. We engineer the following temporal features specific to the surgical operation:

- operation length
- primary procedure operation length percentile
- deviation from the median primary procedure operation length
- operation end hour of day

7.2.2 Prediction Performance

Table 7.2 shows the test set performance of the nine machine learning models and their voting ensemble. Note that we do not have surgeon's prediction for post-operative LOS, so we only compare the models' performance. The voting ensemble model achieves the best result, with an accuracy of 70.1% and 87.0% with one-day error tolerance. Random forest, bagging and XGBoost are competitive models that achieve a comparable performance in MAE and accuracy.

Table 7.2: Post-operation LOS Prediction Performance

Model	**Accuracy**	**Accuracy ($\epsilon = 1$)**	**MAE**	**Overprediction**	**Underprediction**
LGR	67.9%	86.7%	0.56	11.6%	20.5%
SVM	67.6%	85.5%	0.61	10.2%	22.2%
KNN	67.3%	85.3%	0.60	8.5%	24.3%
DT	68.4%	86.0%	0.58	11.3%	20.2%
RF	69.6%	87.1%	0.55	10.5%	19.9%
ExtraTrees	69.3%	86.9%	0.55	10.2%	20.6%
BAG-DT	69.7%	87.3%	0.54	10.2%	20.0%
BST-ADA	67.7%	86.6%	0.58	12.7%	19.6%
XGB	69.5%	87.2%	0.54	11.1%	19.3%
Voting Ensemble	**70.1%**	**87.0%**	**0.54**	**9.3%**	**20.6%**

Confusion Matrix

Fig 7.3 shows the confusion matrix of the voting ensemble model in post-operative LOS prediction. The model shows a strong pattern of underprediction, predicting class 1, 3, 6 with highest frequency and almost never predicts class 5.

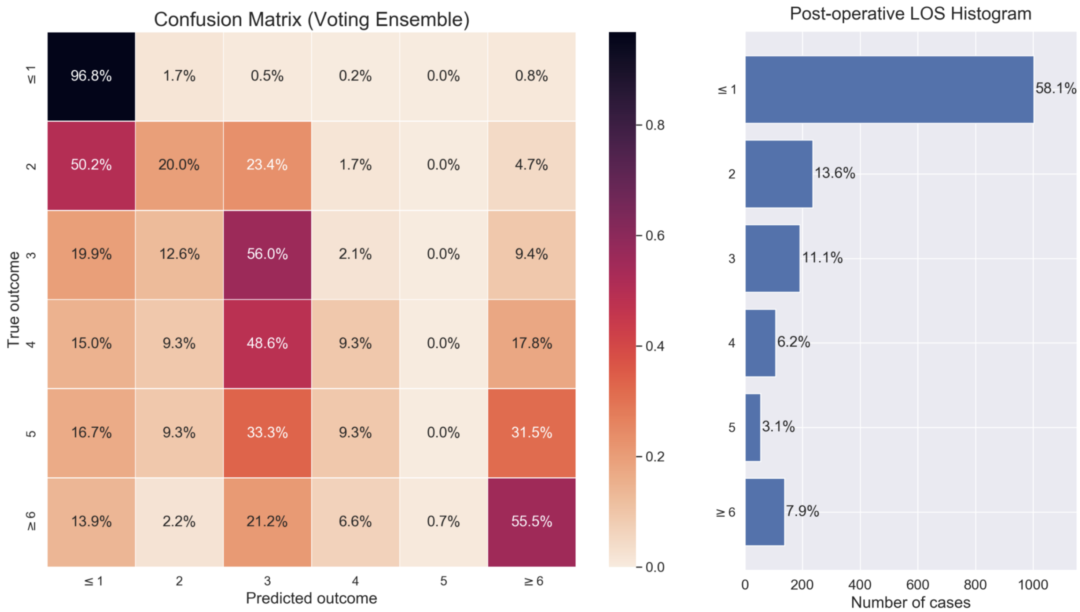


Figure 7.3: Confusion matrix (left) of voting ensemble model in post-operative LOS prediction and the post-operative LOS histogram (right). In the plot, each row is the true LOS outcome class, and each column is the model-predicted outcome class. Each grid is normalized by the true outcome class (row sum). Darker grid implies more data points fall under the corresponding combination of true and predicted class. The right plot is the histogram of LOS distribution over the six outcome classes.

Feature Importance

Fig 7.4 is the ranking of 15 most important features by the voting ensemble in post-operative LOS prediction. Similar to on-admission prediction, care class and procedure-relevant features are ranked as the most important features, next to which are the operative length features. Operation length percentile (within the primary procedure cohort) and the raw operative length are indicative of very short and very long LOS but are not very predictive of the middle ranges. The deviation from the median primary procedure operation length shows the same predictive power, but is overall a weaker predictor compared to other factors.

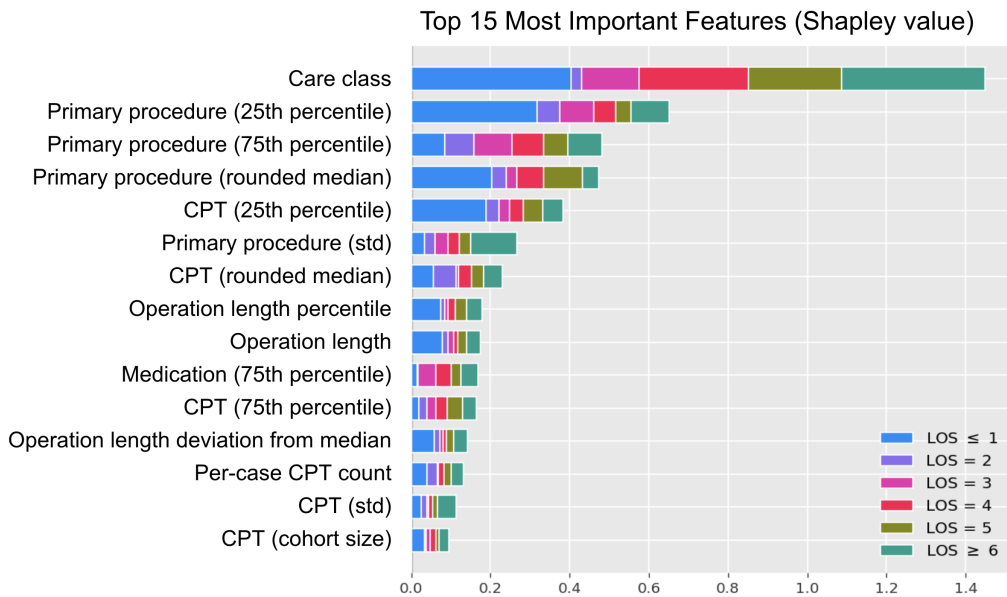


Figure 7.4: Top 15 most important features by voting ensemble model in post-operative LOS prediction. $x-$axis is the Shapley value (larger means more important), and $y-$axis is the feature name. Each bar is the overall Shapley value where the color indicates the contribution of the each feature to predicting the corresponding outcome class.

Chapter 8

Conclusion

This study takes a data-driven approach to infer patient hospital length of stay based on their demographics, diagnosis, surgical procedure and medication characteristics. Without hand-selection of important variables by clinical experts, a well-tuned machine learning model demonstrates human-level intelligence in predicting LOS at scheduling time. Surgeon's prediction and the clinical categorization are helpful to quantify the prediction uncertainty. Adding information available on and after patient admission improves the predictive performance. However, there are important limitations which future work needs to address.

8.1 Limitations

The limitation of this study is three-fold. First, the feature set contains only static information with limited representation power. For example, we do not include vital sign information or lab test results, which arrive dynamically throughout patient time course and correlate with the real-time patient health status. Moreover, the indicator features of the clinical categories only provide the exist-or-not information without direct implication of their medical complexity, thus limited in their predictive power. In addition, the intrinsic noise in the dataset, such as incomplete or missing health records, potentially leads to misinterpretation of the relevant features in the training of the machine learning models.

Second, this study adopts a holistic approach to predict LOS for all clinical cohorts, and ranks the feature importance based on the respective marginal contribution to the final prediction. The importance of different feature combinations is unexplored due to the exponential complexity in the discrete feature combination and the data sparsity within each combination. Therefore, it is imperative to develop a systematic modeling approach that accounts for feature combinations.

Third, the generalizability of this study needs further testing with respect to the patient population as well as the time period. This study considers only patients admitted to Boston Children's Hospital, which does not necessarily have the same characteristics as patients from other hospitals. Whether the models could "transfer" learn from a different population needs further research. In addition, this is a retrospective study which spans a period since the start of Covid-19, and our test set spans the Omicron wave. More data points need to be collected to verify the strength of the machine learning models.

8.2 Future Work

The future work lies in several aspects. To address the first limitation, we need to experiment with different representations of the clinical variables, jointly with more sophisticated modeling approaches. For instance, the autoencoder framework could be used to encode the categorical clinical variables with efficient dimensionality reduction, and deep neural networks could be investigated as an alternative to the "shallow" models discussed in this study. Furthermore, it is important to experiment with different ensemble techniques, such as selecting the optimal model for each clinical cohort, to fine tune the prediction performance. At last, it is crucial to gather feedback from surgeons as end users of our machine learning approach, to jointly contribute to developing a smart decision support system.

Appendix A

Clinical Variable Details

A.1 CCSR Diagnosis

A.2 Primary Procedure

Table A.1: Summary statistics of CCSR diagnosis

	Training data (n=19281)	Test data (n=1777)		Training data (n=19281)	Test data (n=1777)
CCSR, no. %					
Abdominal wall congenital malformation, exomphalos	76 (0.4%)	12 (0.7%)	Abdominal wall congenital malformation, gastroschisis	107 (0.6%)	4 (0.2%)
Abdominal wall congenital malformation, other	6 (0.0%)	0 (0.0%)	Abdominal wall congenital malformation, prune belly	15 (0.1%)	1 (0.1%)
Absence, limb	88 (0.5%)	13 (0.7%)	Absence, other musculoskeletal	1 (0.0%)	0 (0.0%)
Achondroplasia	16 (0.1%)	1 (0.1%)	Adrenal dysfunction, other	64 (0.3%)	6 (0.3%)
Adrenal insufficiency	342 (1.8%)	32 (1.8%)	Adrenal, congenital malformation	5 (0.0%)	1 (0.1%)
Agranulocytosis, congenital	11 (0.1%)	0 (0.0%)	Agranulocytosis, other	12 (0.1%)	3 (0.2%)
Allergic bronchopulmonary aspergillosis	18 (0.1%)	0 (0.0%)	Alpha-1-antitrypsin deficiency	4 (0.0%)	1 (0.1%)
Alport syndrome	24 (0.1%)	1 (0.1%)	Ambulation disorder	277 (1.4%)	21 (1.2%)
Amino acid disorder, cystine	9 (0.0%)	2 (0.1%)	Amino acid disorder, glycine	4 (0.0%)	1 (0.1%)
Amino acid disorder, homocystinuria	9 (0.0%)	0 (0.0%)	Amino acid disorder, hyperphenylalainemia	1 (0.0%)	3 (0.2%)
Amino acid disorder, lysine	9 (0.0%)	0 (0.0%)	Amino acid disorder, other	25 (0.1%)	2 (0.1%)
Amino acid disorder, phenylketonuria	3 (0.0%)	0 (0.0%)	Amino acid disorder, tryptophan	1 (0.0%)	0 (0.0%)
Amino acid disorder, tyrosine	3 (0.0%)	0 (0.0%)	Amyloidosis	5 (0.0%)	0 (0.0%)
Androgen insensitivity	6 (0.0%)	0 (0.0%)	Anemia, aplastic	403 (2.1%)	45 (2.5%)
Anemia, chronic disease	58 (0.3%)	5 (0.3%)	Anemia, hemolytic	154 (0.8%)	8 (0.5%)
Anemia, iron deficiency	573 (3.0%)	59 (3.3%)	Anemia, nutritional	74 (0.4%)	16 (0.9%)
Anemia, sickle cell	96 (0.5%)	10 (0.6%)	Aneurism, aorta	350 (1.8%)	49 (2.8%)
Aneurism, cerebral	7 (0.0%)	0 (0.0%)	Aneurism, heart	27 (0.1%)	3 (0.2%)
Aneurism, other	14 (0.1%)	2 (0.1%)	Angelman syndrome	6 (0.0%)	3 (0.2%)
Anhidrosis	2 (0.0%)	0 (0.0%)	Ankylosing spondylitis	17 (0.1%)	1 (0.1%)
Ankylosis, joint	26 (0.1%)	2 (0.1%)	Anomalous portal venous connection	5 (0.0%)	1 (0.1%)
Anomalous pulmonary venous connection	50 (0.3%)	9 (0.5%)	Anopthalmos	8 (0.0%)	0 (0.0%)
Anorectal congenital malformation	421 (2.2%)	55 (3.1%)	Anterior segment congenital malformation	9 (0.0%)	0 (0.0%)
Antiphospholipid antibody or syndrome	8 (0.0%)	0 (0.0%)	Anxiety and fear-related disorders	1461 (7.6%)	158 (8.9%)
Aorta, coarctation	143 (0.7%)	15 (0.8%)	Aorta, congenital malformation other	273 (1.4%)	19 (1.1%)
Aortic and peripheral arterial embolism or thrombosis	43 (0.2%)	3 (0.2%)	Aortopulmonary septal defect	8 (0.0%)	0 (0.0%)
Aphasia	34 (0.2%)	2 (0.1%)	Arnold-Chiari syndrome	301 (1.6%)	27 (1.5%)
Arterial dissections	4 (0.0%)	1 (0.1%)	Arterial tortuosity syndrome	6 (0.0%)	0 (0.0%)
Arteriovenous fistula	20 (0.1%)	0 (0.0%)	Arteriovenous malformation	161 (0.8%)	7 (0.4%)
Arteritis, Takayasu	10 (0.1%)	1 (0.1%)	Arteritis, giant cell	10 (0.1%)	0 (0.0%)
Arteritis, other	19 (0.1%)	1 (0.1%)	Arteritis, poly	2 (0.0%)	0 (0.0%)
Arthritis, juvenile other	36 (0.2%)	3 (0.2%)	Arthritis, juvenile rheumatoid	53 (0.3%)	2 (0.1%)
Arthritis, other	37 (0.2%)	3 (0.2%)	Arthritis, rheumatoid	19 (0.1%)	0 (0.0%)
Arthrogryposis	54 (0.3%)	6 (0.3%)	Arthropathy, crystal gout	4 (0.0%)	1 (0.1%)
Arthropathy, crystal other	3 (0.0%)	0 (0.0%)	Arthropathy, enteropathic	2 (0.0%)	0 (0.0%)
Arthropathy, other	48 (0.2%)	5 (0.3%)	Arthropathy, traumatic	220 (1.1%)	20 (1.1%)
Articular cartilage disorder	155 (0.8%)	10 (0.6%)	Artificial eye	2 (0.0%)	0 (0.0%)
Artificial joint	65 (0.3%)	6 (0.3%)	Artificial larynx	2 (0.0%)	0 (0.0%)
Artificial limb	8 (0.0%)	0 (0.0%)	Aseptic necrosis and osteonecrosis	126 (0.7%)	8 (0.5%)
Asplenia	26 (0.1%)	2 (0.1%)	Asthma	2194 (11.4%)	165 (9.3%)
Ataxia	17 (0.1%)	5 (0.3%)	Atherosclerosis, coronary and other heart disease	11 (0.1%)	2 (0.1%)
Atherosclerosis, peripheral and visceral	60 (0.3%)	8 (0.5%)	Atrial septal defect	1262 (6.5%)	120 (6.8%)
Atrioventricular septal defect	200 (1.0%)	15 (0.8%)	Attention deficit and Attention deficit hyperactivity disorder	760 (3.9%)	85 (4.8%)
Autism spectrum disorder	697 (3.6%)	68 (3.8%)	Autoimmune lymphoproliferative syndrome [ALPS]	4 (0.0%)	0 (0.0%)
Autoinflammatory syndromes	17 (0.1%)	1 (0.1%)	Autonomic nervous system disorder	171 (0.9%)	22 (1.2%)
Barret's esophagus	15 (0.1%)	1 (0.1%)	Benign intracranial hypertension	64 (0.3%)	6 (0.3%)
Besnier's prurigo	1 (0.0%)	0 (0.0%)	Bile acid and cholesterol metabolism disorder	2 (0.0%)	0 (0.0%)
Biliary atresia	23 (0.1%)	2 (0.1%)	Biliary tract disease, other	173 (0.9%)	7 (0.4%)
Bilirubin metabolism disorder	177 (0.9%)	44 (2.5%)	Biotinidase deficiency	2 (0.0%)	0 (0.0%)
Bipolar and related disorders	42 (0.2%)	4 (0.2%)	Birth trauma, neurologic	2 (0.0%)	1 (0.1%)
Bladder congenital malformation	1 (0.0%)	0 (0.0%)	Bladder disorder, other	210 (1.1%)	23 (1.3%)
Bladder diverticulum	81 (0.4%)	11 (0.6%)	Bladder extrophy	117 (0.6%)	10 (0.6%)
Blindness and low vision	548 (2.8%)	41 (2.3%)	Bowel incontinence	85 (0.4%)	7 (0.4%)
Brachial cleft	27 (0.1%)	4 (0.2%)	Brain congential malformation, other	709 (3.7%)	71 (4.0%)
Breast congenital malformation	46 (0.2%)	5 (0.3%)	Brown-Sequard syndrome	1 (0.0%)	0 (0.0%)
Bullous skin disorder, other	1 (0.0%)	0 (0.0%)	Calculus of bile duct	2 (0.0%)	1 (0.1%)
Cancer, brain and nervous system	306 (1.6%)	31 (1.7%)	Cancer, cardiac	19 (0.1%)	2 (0.1%)
Cancer, endcocrine system - thyroid	37 (0.2%)	5 (0.3%)	Cancer, endocrine system - adrenal	93 (0.5%)	15 (0.8%)
Cancer, endocrine system - other	57 (0.3%)	12 (0.7%)	Cancer, eye	37 (0.2%)	2 (0.1%)
Cancer, female reproductive system - cervix	3 (0.0%)	0 (0.0%)	Cancer, female reproductive system - other	1 (0.0%)	0 (0.0%)
Cancer, female reproductive system - ovary	15 (0.1%)	2 (0.1%)	Cancer, gastrointestinal - colorectal	5 (0.0%)	0 (0.0%)
Cancer, gastrointestinal - liver and biliary	51 (0.3%)	8 (0.5%)	Cancer, gastrointestinal - other	26 (0.1%)	6 (0.3%)
Cancer, gastrointestinal - pancreas	3 (0.0%)	1 (0.1%)	Cancer, gastrointestinal - peritoneum	11 (0.1%)	0 (0.0%)
Cancer, gastrointestinal - small intestine	1 (0.0%)	1 (0.1%)	Cancer, gastrointestinal - stomach	9 (0.0%)	1 (0.1%)
Cancer, head and neck	36 (0.2%)	5 (0.3%)	Cancer, male reproductive - prostate	6 (0.0%)	0 (0.0%)
Cancer, male reproductive - testis	13 (0.1%)	0 (0.0%)	Cancer, other	315 (1.6%)	35 (2.0%)
Cancer, respiratory	10 (0.1%)	2 (0.1%)	Cancer, sarcoma	159 (0.8%)	22 (1.2%)
Cancer, skin - melanoma	6 (0.0%)	0 (0.0%)	Cancer, urinary system - bladder	237 (1.2%)	26 (1.5%)
Cancer, urinary system - kidney	48 (0.2%)	5 (0.3%)	Cancer, urinary system - other	1 (0.0%)	0 (0.0%)
Carbohydrate metabolism disorder, fructose	2 (0.0%)	0 (0.0%)	Carbohydrate metabolism disorder, galactose	1 (0.0%)	1 (0.1%)
Carbohydrate metabolism disorder, glycogen storage	31 (0.2%)	3 (0.2%)	Carbohydrate metabolism disorder, other	11 (0.1%)	0 (0.0%)
Carbohydrate metabolism disorder, pyruvate	15 (0.1%)	1 (0.1%)	Carbohydrate metabolism disorder, sucrase	3 (0.0%)	0 (0.0%)
Cardiac and circulatory congenital malformation - other	182 (0.9%)	10 (0.6%)	Cardiac arrest due to other underlying condition	8 (0.0%)	0 (0.0%)
Cardiac congenital malformation, other	688 (3.6%)	67 (3.8%)	Cardiac septum congenital malformation	22 (0.1%)	2 (0.1%)
Cardiac valve disorder	2475 (12.8%)	257 (14.5%)	Cardiomegaly	725 (3.8%)	78 (4.4%)
Cardiomyopathy	145 (0.8%)	19 (1.1%)	Carpal tunnel syndrome	11 (0.1%)	1 (0.1%)
Cataract	126 (0.7%)	7 (0.4%)	Cauda equina	17 (0.1%)	0 (0.0%)
Celiac disease	79 (0.4%)	9 (0.5%)	Cerebral infarction	147 (0.8%)	19 (1.1%)
Cerebral ischemia attack	30 (0.2%)	3 (0.2%)	Cerebral ischemia syndrome	24 (0.1%)	2 (0.1%)
Cerebral palsy	1043 (5.4%)	85 (4.8%)	Cerebrovascular disease, late effects	111 (0.6%)	10 (0.6%)
Cerebrovascular disease, occlusion and stenosis	55 (0.3%)	8 (0.5%)	Cerebrovascular disease, other	55 (0.3%)	5 (0.3%)
Charcot's arthropathy	1 (0.0%)	0 (0.0%)	Charcot's joint	2 (0.0%)	0 (0.0%)
Chest congenital malformation, other	95 (0.5%)	7 (0.4%)	Choanal atresia	96 (0.5%)	5 (0.3%)
Cholangitis	36 (0.2%)	2 (0.1%)	Chorea	46 (0.2%)	5 (0.3%)
Chromosome abnormalities, other	542 (2.8%)	48 (2.7%)	Chromosome deletion	182 (0.9%)	21 (1.2%)
Chromosome duplication	12 (0.1%)	1 (0.1%)	Chromosome rearrangement	2 (0.0%)	0 (0.0%)
Chromosome ring	6 (0.0%)	0 (0.0%)	Chronic allergic conjunctivitis	73 (0.4%)	9 (0.5%)
Chronic bronchitis	124 (0.6%)	2 (0.1%)	Chronic cystitis	8 (0.0%)	0 (0.0%)
Chronic esophageal problem, other	313 (1.6%)	18 (1.0%)	Chronic eustachian salpingitis	13 (0.1%)	1 (0.1%)
Chronic fatigue syndrome	67 (0.3%)	9 (0.5%)	Chronic gastritis	124 (0.6%)	6 (0.3%)

Table A.1: Summary statistics of CCSR diagnosis variables

	Training data (n=19281)	Test data (n=1777)		Training data (n=19281)	Test data (n=1777)
hspace3mmChronic gingivitis	57 (0.3%)	5 (0.3%)	Chronic headache, migraine	497 (2.6%)	39 (2.2%)
Chronic headache, other	143 (0.7%)	20 (1.1%)	Chronic hepatic failure	5 (0.0%)	0 (0.0%)
Chronic hepatitis	62 (0.3%)	5 (0.3%)	Chronic hypotension	1 (0.0%)	0 (0.0%)
Chronic infection, conjunctivitis	8 (0.0%)	0 (0.0%)	Chronic infection, layrngitis	5 (0.0%)	1 (0.1%)
Chronic infection, nervous system and other	33 (0.2%)	7 (0.4%)	Chronic infection, osteomyelitis	47 (0.2%)	3 (0.2%)
Chronic infection, otitis externa	13 (0.1%)	1 (0.1%)	Chronic infection, otitis media	831 (4.3%)	45 (2.5%)
Chronic infection, pelvic inflammatory disease	2 (0.0%)	0 (0.0%)	Chronic infection, pharyngitis	19 (0.1%)	1 (0.1%)
Chronic infection, sinusitis	318 (1.6%)	11 (0.6%)	Chronic infection, tonsils and adenoids	146 (0.8%)	4 (0.2%)
Chronic infection, urinary tract	1 (0.0%)	0 (0.0%)	Chronic intestine disorder, other	16 (0.1%)	0 (0.0%)
Chronic kidney disease	495 (2.6%)	57 (3.2%)	Chronic lung disease, external agent	2 (0.0%)	0 (0.0%)
Chronic lymphatic disoder	105 (0.5%)	10 (0.6%)	Chronic obstructive pulmonary disease and bronchiectasis	301 (1.6%)	14 (0.8%)
Chronic pain	489 (2.5%)	74 (4.2%)	Chronic pancreatitis	44 (0.2%)	4 (0.2%)
Chronic pericarditis and pericardial disease	1 (0.0%)	0 (0.0%)	Chronic periodontitis	3 (0.0%)	0 (0.0%)
Chronic phlebitis; thrombophlebitis and thromboembolism	50 (0.3%)	2 (0.1%)	Chronic respiratory disease, other	62 (0.3%)	4 (0.2%)
Chronic respiratory insufficiency	795 (4.1%)	87 (4.9%)	Chronic rheumatic heart disease	52 (0.3%)	3 (0.2%)
Chronic rhinitis	937 (4.9%)	71 (4.0%)	Chronic scleral disease	2 (0.0%)	0 (0.0%)
Chronic small intestine problem	14 (0.1%)	0 (0.0%)	Chronic splenomegaly	7 (0.0%)	1 (0.1%)
Chronic thrombocytopenia	4 (0.0%)	0 (0.0%)	Chronic thyroiditis	79 (0.4%)	7 (0.4%)
Chronic ulcer, skin	219 (1.1%)	13 (0.7%)	Chronic uveitis and ocular inflammation	6 (0.0%)	1 (0.1%)
Chronic venous hypertension	1 (0.0%)	0 (0.0%)	Cleft lip	688 (3.6%)	40 (2.3%)
Cleft palate	420 (2.2%)	34 (1.9%)	Coagulation and hemorrhagic disorders, other	98 (0.5%)	4 (0.2%)
Combined immunodeficiency	32 (0.2%)	4 (0.2%)	Common variable immunodeficiency	31 (0.2%)	5 (0.3%)
Communication disorder	49 (0.3%)	4 (0.2%)	Complement system defect	9 (0.0%)	1 (0.1%)
Complete loss of teeth	12 (0.1%)	0 (0.0%)	Complication of other surgical or medical care, injury, initial encounter	1 (0.0%)	0 (0.0%)
Conditions due to neoplasm or the treatment of neoplasm	4 (0.0%)	0 (0.0%)	Congenital malformation, other	1015 (5.3%)	95 (5.3%)
Congenital musculoskeletal malformation, overgrowth syndrome	102 (0.5%)	6 (0.3%)	Congenital musculoskeletal malformation, short stature syndrome	86 (0.4%)	12 (0.7%)
Conjoined twins	8 (0.0%)	0 (0.0%)	Connective tissue disorder, other	66 (0.3%)	2 (0.1%)
Constipation	558 (2.9%)	68 (3.8%)	Cor triatriatum	1 (0.0%)	0 (0.0%)
Cornea and external disease, other	6 (0.0%)	0 (0.0%)	Corneal congenital malformation	18 (0.1%)	2 (0.1%)
Cyclic vomiting syndrome	15 (0.1%)	5 (0.3%)	Cystic fibrosis	159 (0.8%)	2 (0.1%)
Cystic kidney disease	133 (0.7%)	17 (1.0%)	Cystostomy	113 (0.6%)	7 (0.4%)
Delirium	22 (0.1%)	2 (0.1%)	Dementia	3 (0.0%)	1 (0.1%)
Demyelinating disease of central nervous system	18 (0.1%)	3 (0.2%)	Dependence on enabling machine or device	154 (0.8%)	38 (2.1%)
Depressive disorders	648 (3.4%)	58 (3.3%)	Dermatitis, atopic	391 (2.0%)	36 (2.0%)
Dermatitis, neuro	1 (0.0%)	0 (0.0%)	Dermatitis, perioral	23 (0.1%)	2 (0.1%)
Dermatomyositis	10 (0.1%)	1 (0.1%)	Dextrocardia	42 (0.2%)	2 (0.1%)
Di George's syndrome	51 (0.3%)	9 (0.5%)	Diabetes insipidus	52 (0.3%)	3 (0.2%)
Diabetes mellitus, Type I	52 (0.3%)	6 (0.3%)	Diabetes mellitus, Type II	65 (0.3%)	4 (0.2%)
Diabetes mellitus, neonatal	3 (0.0%)	2 (0.1%)	Diabetes mellitus, other	65 (0.3%)	2 (0.1%)
Diaphragmatic hernia	78 (0.4%)	9 (0.5%)	Diaphram congenital malformation, other	32 (0.2%)	5 (0.3%)
Digestive congenital malformation, lower	274 (1.4%)	13 (0.7%)	Digestive congenital malformation, other	20 (0.1%)	4 (0.2%)
Digestive congenital malformation, upper	7 (0.0%)	0 (0.0%)	Discordand ventricular connection, other	49 (0.3%)	7 (0.4%)
Diseases of white blood cells	46 (0.2%)	7 (0.4%)	Disruptive, impulse-control and conduct disorders	337 (1.7%)	15 (0.8%)
Diverticulitis and Diverticulosis	15 (0.1%)	1 (0.1%)	Dolichocephaly	29 (0.2%)	3 (0.2%)
Double inlet ventricle	36 (0.2%)	4 (0.2%)	Double outlet right ventricle	62 (0.3%)	5 (0.3%)
Dyspareunia	11 (0.1%)	0 (0.0%)	Dysrhythmia, atrioventicular heart block	81 (0.4%)	10 (0.6%)
Dysrhythmia, chronic atrial fibrillation	7 (0.0%)	0 (0.0%)	Dysrhythmia, other	1832 (9.5%)	167 (9.4%)
Dysrhythmia, other heart block	268 (1.4%)	22 (1.2%)	Dystonia	444 (2.3%)	41 (2.3%)
Ear congenital malformation, other	250 (1.3%)	16 (0.9%)	Ebstein's anomaly	11 (0.1%)	0 (0.0%)
Ectodermal dysplasia	12 (0.1%)	2 (0.1%)	Ehlers-Danlos syndrome	134 (0.7%)	22 (1.2%)
Emphysema	16 (0.1%)	1 (0.1%)	Encephalitis, other	4 (0.0%)	0 (0.0%)
Encephalocele	27 (0.1%)	2 (0.1%)	Encephalopathy, hypoxic ischemic	170 (0.9%)	17 (1.0%)
Encephalopathy, other	272 (1.4%)	11 (0.6%)	Encounter for antineoplastic therapies	257 (1.3%)	33 (1.9%)
Endocarditis and endocardial disease	10 (0.1%)	0 (0.0%)	Endocrine disorder, other	23 (0.1%)	2 (0.1%)
Endocrine gland, congenital malformation	68 (0.4%)	5 (0.3%)	Endometriosis	58 (0.3%)	10 (0.6%)
Enopthalamos	3 (0.0%)	0 (0.0%)	Enterostomy	2308 (12.0%)	210 (11.8%)
Epidermolysis bullosa	1 (0.0%)	0 (0.0%)	Epilepsy	1379 (7.2%)	123 (6.9%)
Epispadias	29 (0.2%)	2 (0.1%)	Esophageal atresia/tracheoesophageal fistula	469 (2.4%)	36 (2.0%)
Esophageal reflux	2797 (14.5%)	250 (14.1%)	Esophageal stenosis or stricture	197 (1.0%)	17 (1.0%)
Esophageal varices	45 (0.2%)	1 (0.1%)	Esophageal web	4 (0.0%)	0 (0.0%)
Esophagitis	139 (0.7%)	24 (1.4%)	Esophagus congenital malformation	37 (0.2%)	1 (0.1%)
Exophthalmos	29 (0.2%)	2 (0.1%)	Extremity congenital malformation	1259 (6.5%)	112 (6.3%)
Eye congenital malformation, other	27 (0.1%)	4 (0.2%)	Eye hypotony	1 (0.0%)	0 (0.0%)
Eye vascular disorder, other	1 (0.0%)	0 (0.0%)	Eyelid congenital malformation	95 (0.5%)	17 (1.0%)
Face and skull congenital malformation, asymmetry	2 (0.0%)	0 (0.0%)	Face and skull congenital malformation, compression	128 (0.7%)	10 (0.6%)
Face and skull congenital malformation, craniosynostosis	382 (2.0%)	19 (1.1%)	Face and skull congenital malformation, dysostosis	39 (0.2%)	3 (0.2%)
Face and skull congenital malformation, other	681 (3.5%)	63 (3.5%)	Factor IX deficiency	4 (0.0%)	0 (0.0%)
Factor VIII deficiency	11 (0.1%)	0 (0.0%)	Factor XI deficiency	4 (0.0%)	0 (0.0%)
Fatty acid metabolism disorder, carnitine	12 (0.1%)	1 (0.1%)	Fatty acid metabolism disorder, glutaric aciduria type 2	2 (0.0%)	0 (0.0%)
Fatty acid metabolism disorder, long	1 (0.0%)	0 (0.0%)	Fatty acid metabolism disorder, medium	7 (0.0%)	0 (0.0%)
Fatty acid metabolism disorder, other	7 (0.0%)	0 (0.0%)	Fatty acid metabolism disorder, short	2 (0.0%)	0 (0.0%)
Feeding and eating disorders	535 (2.8%)	68 (3.8%)	Fetal alcohol syndrome	11 (0.1%)	1 (0.1%)
Foot adquired deformity	97 (0.5%)	9 (0.5%)	Foot congenital malformation	675 (3.5%)	61 (3.4%)
Fragile X syndrome	2 (0.0%)	0 (0.0%)	Frontal lobe and executive function deficit	6 (0.0%)	1 (0.1%)
Gender identity disorder	304 (1.6%)	38 (2.1%)	General paresis	7 (0.0%)	0 (0.0%)
Glaucoma	60 (0.3%)	5 (0.3%)	Globe disorder	7 (0.0%)	0 (0.0%)
Glomerular abnormality, diffuse	15 (0.1%)	0 (0.0%)	Glomerular abnormality, focal and segmental	21 (0.1%)	1 (0.1%)
Glomerular abnormality, other	74 (0.4%)	1 (0.1%)	Glucosaminoglycan disorder	15 (0.1%)	3 (0.2%)
Graft-versus-host disease	24 (0.1%)	3 (0.2%)	Great vessel, congenital malformation other	537 (2.8%)	62 (3.5%)
Guillian-Barre syndrome	1 (0.0%)	0 (0.0%)	HIV infection	7 (0.0%)	0 (0.0%)
Hand congenital malformation	286 (1.5%)	17 (1.0%)	Hearing loss	2607 (13.5%)	221 (12.4%)
Heart disease, other	116 (0.6%)	16 (0.9%)	Heart failure	427 (2.2%)	42 (2.4%)
Heavy metal toxic effect	3 (0.0%)	0 (0.0%)	Hematologic disorder, other	38 (0.2%)	3 (0.2%)
Hemiplegia	305 (1.6%)	28 (1.6%)	Hemochromatosis	34 (0.2%)	6 (0.3%)
Hemophilia	4 (0.0%)	0 (0.0%)	Hepatic fibrosis	22 (0.1%)	0 (0.0%)
Hepatopulmonary syndrome	7 (0.0%)	1 (0.1%)	Hepatorenal syndrome	1 (0.0%)	0 (0.0%)
Hereditary hemorrhagic telangiectasia	4 (0.0%)	1 (0.1%)	Hermansky-Pudlak syndrome	2 (0.0%)	0 (0.0%)
Hermaphroditism	72 (0.4%)	7 (0.4%)	Herpes	3 (0.0%)	1 (0.1%)
Hiatal hernia	1 (0.0%)	0 (0.0%)	Hip, congenital deformity	1826 (9.5%)	176 (9.9%)
Hirschsprung's disease	92 (0.5%)	13 (0.7%)	Histiocytosis syndrome	18 (0.1%)	2 (0.1%)

Table A.1: Summary statistics of CCSR diagnosis variables

	Training data (n=19281)	Test data (n=1777)		Training data (n=19281)	Test data (n=1777)
Holoprosencephaly	22 (0.1%)	1 (0.1%)	Horseshoe kidney	68 (0.4%)	10 (0.6%)
Hungry bone syndrome	3 (0.0%)	0 (0.0%)	Hydrocephalus	710 (3.7%)	67 (3.8%)
Hydronephrosis	411 (2.1%)	54 (3.0%)	Hyperaldosteronism	118 (0.6%)	10 (0.6%)
Hyperalimentation	24 (0.1%)	3 (0.2%)	Hypercholesterolemia	157 (0.8%)	15 (0.8%)
Hypergammaglobulinemia	23 (0.1%)	0 (0.0%)	Hyperimmunoglobulin E syndrome	3 (0.0%)	0 (0.0%)
Hyperlipidemia	120 (0.6%)	14 (0.8%)	Hyperparathyroidism	216 (1.1%)	30 (1.7%)
Hypertension	761 (3.9%)	64 (3.6%)	Hyperthyroidism	46 (0.2%)	7 (0.4%)
Hypertriglyceridemia, lipoprotein deficiency	33 (0.2%)	3 (0.2%)	Hypogammaglobulinemia	283 (1.5%)	27 (1.5%)
Hypoparathyroidism	33 (0.2%)	6 (0.3%)	Hypophosphatemia, familial	19 (0.1%)	0 (0.0%)
Hypopituitarism	168 (0.9%)	19 (1.1%)	Hypoplastic left heart syndrome	63 (0.3%)	5 (0.3%)
Hypoplastic right heart syndrome	7 (0.0%)	1 (0.1%)	Hypospadius	296 (1.5%)	31 (1.7%)
Hypothyroidism	501 (2.6%)	41 (2.3%)	Ichthyosis	30 (0.2%)	1 (0.1%)
Immunodeficiency antibody, other	7 (0.0%)	1 (0.1%)	Immunodeficiency, other	877 (4.5%)	102 (5.7%)
Implant, endocrine	2 (0.0%)	0 (0.0%)	Implant, urogenital	59 (0.3%)	5 (0.3%)
Indwelling medical device, audiological implant	91 (0.5%)	10 (0.6%)	Indwelling medical device, cardiac other	552 (2.9%)	43 (2.4%)
Indwelling medical device, cardiac pacemaker	51 (0.3%)	8 (0.5%)	Indwelling medical device, cerebrospinal fluid device	389 (2.0%)	42 (2.4%)
Indwelling medical device, debribrillator	10 (0.1%)	2 (0.1%)	Indwelling medical device, dialysis	127 (0.7%)	10 (0.6%)
Indwelling medical device, digestive	138 (0.7%)	20 (1.1%)	Indwelling medical device, insulin pump	47 (0.2%)	5 (0.3%)
Indwelling medical device, myringotomy device	689 (3.6%)	41 (2.3%)	Indwelling medical device, neurologic other	36 (0.2%)	6 (0.3%)
Indwelling medical device, other	214 (1.1%)	20 (1.1%)	Indwelling medical device, urologic	58 (0.3%)	5 (0.3%)
Indwelling medical device, vascular	379 (2.0%)	50 (2.8%)	Interstitial pulmonary disease	114 (0.6%)	9 (0.5%)
Intervertebral disc disorder	247 (1.3%)	14 (0.8%)	Intestinal bypass and anastamosis	236 (1.2%)	8 (0.5%)
Intracranial hemorrhage, non-traumatic	166 (0.9%)	9 (0.5%)	Iris congenital malformation	39 (0.2%)	8 (0.5%)
Irregular eye movements	48 (0.2%)	4 (0.2%)	Irritable bowel syndrome	131 (0.7%)	16 (0.9%)
Joint contracture, hip	119 (0.6%)	8 (0.5%)	Joint contracture, non-hip	624 (3.2%)	50 (2.8%)
Joint derangement, other	110 (0.6%)	17 (1.0%)	Keratitis	39 (0.2%)	6 (0.3%)
Kidney congenital malformation, other	401 (2.1%)	52 (2.9%)	Kidney disease, other	754 (3.9%)	79 (4.4%)
Klinefelter syndrome	4 (0.0%)	0 (0.0%)	Klippel-Feil syndrome	60 (0.3%)	3 (0.2%)
Knee musculoskeletal disorder	304 (1.6%)	23 (1.3%)	Kyphosis	306 (1.6%)	22 (1.2%)
Lacrimal congenital malformation	61 (0.3%)	6 (0.3%)	Lactase deficiency or intolerance	85 (0.4%)	9 (0.5%)
Large intestine congenital malformation	46 (0.2%)	3 (0.2%)	Larynx congenital malformation	620 (3.2%)	57 (3.2%)
Lens congenital malformation	5 (0.0%)	1 (0.1%)	Lens disorder, other	52 (0.3%)	3 (0.2%)
Lesch-Nyhan syndrome	3 (0.0%)	0 (0.0%)	Leucocoria	5 (0.0%)	0 (0.0%)
Leukemia - acute lymphoblastic leukemia (ALL)	83 (0.4%)	3 (0.2%)	Leukemia - acute myeloid leukemia (AML)	27 (0.1%)	4 (0.2%)
Leukemia - all other types	79 (0.4%)	5 (0.3%)	Leukemia - chronic lymphocytic leukemia (CLL)	1 (0.0%)	0 (0.0%)
Leukemia - chronic myeloid leukemia (CML)	4 (0.0%)	0 (0.0%)	Leukodystrophy	14 (0.1%)	1 (0.1%)
Levocardia	24 (0.1%)	2 (0.1%)	Lipid metabolism disorder	9 (0.0%)	1 (0.1%)
Lipid storage disorder, other	2 (0.0%)	0 (0.0%)	Lipodystrophy	7 (0.0%)	2 (0.1%)
Lipomatosis	9 (0.0%)	1 (0.1%)	Liver and biliary cirrhosis	70 (0.4%)	4 (0.2%)
Liver and biliary congenital malformation	70 (0.4%)	5 (0.3%)	Liver disease, other	413 (2.1%)	38 (2.1%)
Long QT syndrome	195 (1.0%)	21 (1.2%)	Lordosis	108 (0.6%)	12 (0.7%)
Lowe's syndrome	3 (0.0%)	0 (0.0%)	Lung congenital malformation	280 (1.5%)	25 (1.4%)
Lupus erythematosus	10 (0.1%)	3 (0.2%)	Lymphedema	7 (0.0%)	0 (0.0%)
Lymphoma, Hodgkin	8 (0.0%)	1 (0.1%)	Lymphoma, non-Hodgkin	33 (0.2%)	3 (0.2%)
Lysosomal storage disease, Fabry	1 (0.0%)	0 (0.0%)	Lysosomal storage disease, Gangliosidosis	2 (0.0%)	0 (0.0%)
Lysosomal storage disease, Krabbe disease	4 (0.0%)	1 (0.1%)	Lysosomal storage disease, Mucopolysaccharidosis	27 (0.1%)	5 (0.3%)
Lysosomal storage disease, Niemann-Pick	4 (0.0%)	0 (0.0%)	Lysosomal storage disease, Sphingolipidosis	16 (0.1%)	5 (0.3%)
Lysosomal storage disease, Tay-Sachs	2 (0.0%)	1 (0.1%)	Lysosomal storage disease, lipofuscinosis	5 (0.0%)	1 (0.1%)
Lysosomal storage disease, other	3 (0.0%)	2 (0.1%)	MELAS syndrome	1 (0.0%)	0 (0.0%)
Malabsorption, other	398 (2.1%)	29 (1.6%)	Malacia of trachea or larynx	1060 (5.5%)	97 (5.5%)
Malignant neoplasm, unspecified	153 (0.8%)	30 (1.7%)	Malignant neuroendocrine tumors	8 (0.0%)	1 (0.1%)
Malnutrition	934 (4.8%)	116 (6.5%)	Marfan's syndrome	133 (0.7%)	15 (0.8%)
Mast cell activation disorder	10 (0.1%)	2 (0.1%)	Mastocytosis	8 (0.0%)	0 (0.0%)
Meckel's diverticulum	16 (0.1%)	2 (0.1%)	Menopausal disorders	5 (0.0%)	1 (0.1%)
Menstrual disorders	532 (2.8%)	57 (3.2%)	Mental disorder, other	970 (5.0%)	90 (5.1%)
Metabolic disorder, other	297 (1.5%)	23 (1.3%)	Methemaglobinemia	4 (0.0%)	0 (0.0%)
Microephaly	343 (1.8%)	26 (1.5%)	Microtia	56 (0.3%)	5 (0.3%)
Mineral metabolism disorder	562 (2.9%)	76 (4.3%)	Mitochrondrial disorders	48 (0.2%)	4 (0.2%)
Monoarthritis	2 (0.0%)	0 (0.0%)	Mononeuropathy	67 (0.3%)	4 (0.2%)
Monoplegia	46 (0.2%)	6 (0.3%)	Monosomy	1 (0.0%)	2 (0.1%)
Mood disorder, other	53 (0.3%)	8 (0.5%)	Mouth congenital malformation, other	283 (1.5%)	18 (1.0%)
Movement disorder, other	156 (0.8%)	12 (0.7%)	Moyamoya disease	74 (0.4%)	9 (0.5%)
Mucopolysaccharidosis	13 (0.1%)	2 (0.1%)	Multiple endocrine neoplasia	11 (0.1%)	0 (0.0%)
Multiple sclerosis	3 (0.0%)	0 (0.0%)	Muscular dystrophy	63 (0.3%)	4 (0.2%)
Musculoskeletal congenital malformation, other	305 (1.6%)	18 (1.0%)	Myasthenia Gravis	8 (0.0%)	3 (0.2%)
Myocarditis	4 (0.0%)	0 (0.0%)	Myoclonus	59 (0.3%)	3 (0.2%)
Myopathy, orbit	16 (0.1%)	2 (0.1%)	Myopathy, other	82 (0.4%)	5 (0.3%)
Myopathy, rheumatoid	3 (0.0%)	0 (0.0%)	Myopia, degenerative	31 (0.2%)	4 (0.2%)
Narcolepsy	8 (0.0%)	1 (0.1%)	Neonatal abstinence syndrome	8 (0.0%)	1 (0.1%)
Neoplasms of unspecified nature or uncertain behavior	133 (0.7%)	5 (0.3%)	Nephropathy, external agent	10 (0.1%)	3 (0.2%)
Nephrostomy	12 (0.1%)	0 (0.0%)	Nervous system congenital malformation, other	108 (0.6%)	12 (0.7%)
Nervous system disorder, central - other	1442 (7.5%)	104 (5.9%)	Nervous system disorder, peripheral - other	242 (1.3%)	32 (1.8%)
Nervous system disorder, unspecified - other	3 (0.0%)	0 (0.0%)	Neuro-ophthalmologic disorder	427 (2.2%)	37 (2.1%)
Neurodegenerative disease, other	65 (0.3%)	3 (0.2%)	Neurodevelopmental disorder	3343 (17.3%)	277 (15.6%)
Neurofibromatosis	142 (0.7%)	17 (1.0%)	Neurogenic bowel	298 (1.5%)	23 (1.3%)
Neurogenic tic	17 (0.1%)	0 (0.0%)	Neuropathic bladder	694 (3.6%)	64 (3.6%)
Neutropenia, congenital	124 (0.6%)	15 (0.8%)	Neutropenia, cyclic	4 (0.0%)	1 (0.1%)
Nonmalignant breast condition	3 (0.0%)	2 (0.1%)	Nose congenital malformation	63 (0.3%)	4 (0.2%)
Nystagmus	289 (1.5%)	29 (1.6%)	Obesity	928 (4.8%)	84 (4.7%)
Obsessive-compulsive and related disorders	94 (0.5%)	10 (0.6%)	Obstructive genitourinary defect	327 (1.7%)	28 (1.6%)
Oculofacial plastics and orbital conditions	1 (0.0%)	0 (0.0%)	Orbit congenital malformation	26 (0.1%)	4 (0.2%)
Orbit cyst	2 (0.0%)	0 (0.0%)	Orbit deformity, other	3 (0.0%)	0 (0.0%)
Orbital granuloma	1 (0.0%)	0 (0.0%)	Organic acidemia, maple-syrup-urine disease	3 (0.0%)	0 (0.0%)
Organic acidemia, methylmalonic	2 (0.0%)	0 (0.0%)	Organic acidemia, proprionic	3 (0.0%)	1 (0.1%)
Osteoarthritis	118 (0.6%)	9 (0.5%)	Osteochondropathy	36 (0.2%)	5 (0.3%)
Osteochondrosis, hip or pelvis	51 (0.3%)	3 (0.2%)	Osteochondrosis, other joint	85 (0.4%)	2 (0.1%)
Osteochondrosis, spine	6 (0.0%)	0 (0.0%)	Osteochrondritis, dissecans	125 (0.6%)	9 (0.5%)
Osteochrondrodysplasia	114 (0.6%)	11 (0.6%)	Osteodystrophy	6 (0.0%)	0 (0.0%)
Osteogenesis Imperfecta	66 (0.3%)	2 (0.1%)	Osteopathy, hypertrophic	1 (0.0%)	0 (0.0%)
Osteopathy, other	17 (0.1%)	1 (0.1%)	Osteopathy, poliomyelitis	1 (0.0%)	0 (0.0%)

Table A.1: Summary statistics of CCSR diagnosis variables

	Training data (n=19281)	Test data (n=1777)		Training data (n=19281)	Test data (n=1777)
Osteopetrosis	5 (0.0%)	0 (0.0%)	Osteoporosis	148 (0.8%)	12 (0.7%)
Ovary and fallopian tube congenital malformation	8 (0.0%)	3 (0.2%)	Ovary dysfunction	59 (0.3%)	5 (0.3%)
Pancreas congenital malformation	18 (0.1%)	4 (0.2%)	Pancreas disorder, other	86 (0.4%)	1 (0.1%)
Paraplegia and Paralysis	245 (1.3%)	12 (0.7%)	Parathyroid disorder, other	25 (0.1%)	3 (0.2%)
Parkinson's disease	2 (0.0%)	2 (0.1%)	Parkinsonism	7 (0.0%)	0 (0.0%)
Patent ductus arteriosus	497 (2.6%)	49 (2.8%)	Pectus carinatum	48 (0.2%)	5 (0.3%)
Pectus excavatum	129 (0.7%)	6 (0.3%)	Penis congenital malformation	169 (0.9%)	14 (0.8%)
Penis problem, other	3 (0.0%)	0 (0.0%)	Peripheral vascular congenital malformation, other	516 (2.7%)	39 (2.2%)
Peripheral vascular disease, other	217 (1.1%)	28 (1.6%)	Peroxismal disorder	14 (0.1%)	0 (0.0%)
Personality disorder	14 (0.1%)	1 (0.1%)	Phakomatosis	51 (0.3%)	8 (0.5%)
Pituitary disorder, other	129 (0.7%)	8 (0.5%)	Pityriasis	23 (0.1%)	4 (0.2%)
Plagiocephaly	329 (1.7%)	35 (2.0%)	Platelet dysfunction	161 (0.8%)	12 (0.7%)
Polycystic ovarian syndrome	64 (0.3%)	10 (0.6%)	Polycythemia	14 (0.1%)	1 (0.1%)
Polyglandular dysfunction	6 (0.0%)	0 (0.0%)	Polyneuropathy, other	86 (0.4%)	13 (0.7%)
Portal hypertension	66 (0.3%)	3 (0.2%)	Post-transplant lymphoproliferative disorder	14 (0.1%)	3 (0.2%)
Postconcussional syndrome	37 (0.2%)	3 (0.2%)	Posterior segment congenital malformation	99 (0.5%)	11 (0.6%)
Posterior urethral valves	74 (0.4%)	5 (0.3%)	Postthrombic syndrome	3 (0.0%)	0 (0.0%)
Prader-Willi syndrome	10 (0.1%)	4 (0.2%)	Priapism	4 (0.0%)	0 (0.0%)
Protein C resistance	17 (0.1%)	3 (0.2%)	Protrusio acetabuli	13 (0.1%)	0 (0.0%)
Psoriasis	70 (0.4%)	9 (0.5%)	Psychomotor deficit	5 (0.0%)	0 (0.0%)
Puberty, delayed	100 (0.5%)	12 (0.7%)	Puberty, other disorder	47 (0.2%)	0 (0.0%)
Puberty, precocious	192 (1.0%)	17 (1.0%)	Pulmonary heart disease	6 (0.0%)	1 (0.1%)
Pulmonary hypertension	308 (1.6%)	16 (0.9%)	Pulmonary hypertension, newborn	61 (0.3%)	4 (0.2%)
Pupil disorder	81 (0.4%)	11 (0.6%)	Purine and pyrimidine metabolism disorder	10 (0.1%)	0 (0.0%)
Pyloric stenosis	14 (0.1%)	1 (0.1%)	Quadriplegia	291 (1.5%)	15 (0.8%)
Regional enteritis and ulcerative colitis	248 (1.3%)	24 (1.4%)	Renal agenesis	124 (0.6%)	16 (0.9%)
Renal artery stenosis, congenital	23 (0.1%)	3 (0.2%)	Renal dysplasia	239 (1.2%)	28 (1.6%)
Renal hypoplasia	18 (0.1%)	1 (0.1%)	Renal sclerosis	9 (0.0%)	2 (0.1%)
Renal tubulo-interstitial disease	30 (0.2%)	1 (0.1%)	Respiratory congenital malformation, other	40 (0.2%)	3 (0.2%)
Retinal and vitreous conditions	192 (1.0%)	24 (1.4%)	Retinopathy of Prematurity, stage 0 or 1	138 (0.7%)	12 (0.7%)
Retinopathy of Prematurity, stage 2-5	67 (0.3%)	6 (0.3%)	Retinopathy of Prematurity, unspecified	71 (0.4%)	9 (0.5%)
Retinopathy, other	50 (0.3%)	3 (0.2%)	Rhizomelic chondrodysplasia punctata	2 (0.0%)	1 (0.1%)
Rosacea	1 (0.0%)	0 (0.0%)	Ruvalcaba-Myhre-Smith syndrome	3 (0.0%)	0 (0.0%)
Sacroiliitis	25 (0.1%)	3 (0.2%)	Sarcoidosis	2 (0.0%)	1 (0.1%)
Schizophrenia spectrum and other psychotic disorders	18 (0.1%)	2 (0.1%)	Sclera congenital malformation	5 (0.0%)	0 (0.0%)
Scleroderma, localized or linear	10 (0.1%)	0 (0.0%)	Scoliosis, congenital	190 (1.0%)	19 (1.1%)
Scoliosis, ideopathic	1313 (6.8%)	79 (4.4%)	Scoliosis, neuromuscular	627 (3.3%)	71 (4.0%)
Scoliosis, other	1851 (9.6%)	171 (9.6%)	Secondary malignancies	256 (1.3%)	31 (1.7%)
Septo-optic dysplasia	22 (0.1%)	3 (0.2%)	Sequeale of infection, other	28 (0.1%)	2 (0.1%)
Sex chromosome abnormality	45 (0.2%)	4 (0.2%)	Sexual disorder	1 (0.0%)	0 (0.0%)
Sexual identity disorder	28 (0.1%)	12 (0.7%)	Short stature	68 (0.4%)	13 (0.7%)
Sicca syndrome	6 (0.0%)	1 (0.1%)	Situs inversus	28 (0.1%)	1 (0.1%)
Skin congenital malformation	173 (0.9%)	17 (1.0%)	Sleep apnea	3197 (16.6%)	174 (9.8%)
Sleep disorder, other	650 (3.4%)	55 (3.1%)	Slipped upper femoral epiphysis	53 (0.3%)	4 (0.2%)
Small intestine congenital malformation	91 (0.5%)	2 (0.1%)	Smith-Lemli-Opitz syndrome	6 (0.0%)	0 (0.0%)
Somatic disorder	193 (1.0%)	25 (1.4%)	Spina Bifida	442 (2.3%)	37 (2.1%)
Spina Bifida Occulta	40 (0.2%)	4 (0.2%)	Spinal cord congenital malformation, other	677 (3.5%)	76 (4.3%)
Spinal muscular atrophy	53 (0.3%)	6 (0.3%)	Spine, congenital deformity	290 (1.5%)	30 (1.7%)
Spine, dorsopathy	10 (0.1%)	0 (0.0%)	Spleen malformation	24 (0.1%)	0 (0.0%)
Spondylosis and allied disorders	88 (0.5%)	10 (0.6%)	Stomach congenital malformation	18 (0.1%)	1 (0.1%)
Subglottic stenosis	128 (0.7%)	9 (0.5%)	Substance abuse disorder	193 (1.0%)	15 (0.8%)
Suicide and intentional self-inflected injury	3 (0.0%)	0 (0.0%)	Syndrome of inappropriate secretion of antidiuretic hormone	60 (0.3%)	5 (0.3%)
Systemic sclerosis	2 (0.0%)	0 (0.0%)	Testicular dysfunction	48 (0.2%)	5 (0.3%)
Testis congenital malformation	436 (2.3%)	38 (2.1%)	Tetralogy of Fallot	131 (0.7%)	20 (1.1%)
Thrombophilia	35 (0.2%)	0 (0.0%)	Thymus hyperplasia	2 (0.0%)	0 (0.0%)
Thyroid disorder, other	163 (0.8%)	23 (1.3%)	Tic disorder	45 (0.2%)	4 (0.2%)
Tongue malformation	189 (1.0%)	19 (1.1%)	Tonsils or adenoids, hypertrophy	3149 (16.3%)	115 (6.5%)
Tonsils or adenoids, other chronic disease	41 (0.2%)	2 (0.1%)	Torticollis	152 (0.8%)	17 (1.0%)
Tourette's disorder	26 (0.1%)	0 (0.0%)	Trachea congenital malformation	99 (0.5%)	9 (0.5%)
Tracheostomy	515 (2.7%)	39 (2.2%)	Trachoma	1 (0.0%)	0 (0.0%)
Transplant, Bone	2 (0.0%)	1 (0.1%)	Transplant, Bone Marrow	76 (0.4%)	7 (0.4%)
Transplant, Cardiac	36 (0.2%)	7 (0.4%)	Transplant, Cornea	1 (0.0%)	0 (0.0%)
Transplant, Intestine	15 (0.1%)	4 (0.2%)	Transplant, Kidney	122 (0.6%)	18 (1.0%)
Transplant, Liver	82 (0.4%)	7 (0.4%)	Transplant, Lung	36 (0.2%)	1 (0.1%)
Transplant, NOS	209 (1.1%)	24 (1.4%)	Transplant, Pancreas	6 (0.0%)	2 (0.1%)
Transplant, Stem Cell	58 (0.3%)	6 (0.3%)	Trauma and stressor-related mental disorders	847 (4.4%)	64 (3.6%)
Tremor	17 (0.1%)	6 (0.3%)	Triploidy and polyploidy	2 (0.0%)	0 (0.0%)
Trisomy 13	4 (0.0%)	0 (0.0%)	Trisomy 18	17 (0.1%)	2 (0.1%)
Trisomy 21	352 (1.8%)	29 (1.6%)	Trisomy, other	164 (0.9%)	16 (0.9%)
Truncus Arteriosus	9 (0.0%)	1 (0.1%)	Tuberous sclerosis	27 (0.1%)	2 (0.1%)
Tumor Lysis Syndrome	24 (0.1%)	2 (0.1%)	Turner's syndrome	16 (0.1%)	1 (0.1%)
Ulcer, chronic peptic	2 (0.0%)	0 (0.0%)	Ulcer, digestive other	7 (0.0%)	0 (0.0%)
Ulcer, duodenal	35 (0.2%)	5 (0.3%)	Ulcer, esophagus	43 (0.2%)	4 (0.2%)
Ulcer, gastric	56 (0.3%)	8 (0.5%)	Urea cycle disorder	28 (0.1%)	1 (0.1%)
Ureter congenital malformation	204 (1.1%)	22 (1.2%)	Urethral congenital malformation	34 (0.2%)	2 (0.1%)
Urethral disorder, other	17 (0.1%)	1 (0.1%)	Urinary incontinence	667 (3.5%)	69 (3.9%)
Urogenital congenital malformation, other	62 (0.3%)	8 (0.5%)	Urogenital disorder, other	7 (0.0%)	0 (0.0%)
Urogenital fistula	190 (1.0%)	29 (1.6%)	Urogenital prolapse	8 (0.0%)	0 (0.0%)
Uterus congenital malformation	60 (0.3%)	4 (0.2%)	Uveitis and ocular inflammation	6 (0.0%)	0 (0.0%)
Vagina congenital malformation	92 (0.5%)	12 (0.7%)	Vascular disorder intestine	38 (0.2%)	2 (0.1%)
Vasculitis, Bechet's disease	2 (0.0%)	0 (0.0%)	Vasculitis, Wegener's granulomatosis	9 (0.0%)	0 (0.0%)
Vasculitis, rheumatoid	2 (0.0%)	0 (0.0%)	Vasculopathy, other	36 (0.2%)	2 (0.1%)
Velo-cardio-facial syndrome	66 (0.3%)	6 (0.3%)	Ventilator dependence	590 (3.1%)	58 (3.3%)
Ventricular septal defect	499 (2.6%)	50 (2.8%)	Vision loss, other	303 (1.6%)	10 (0.6%)
Vitamin deficiency, D	1155 (6.0%)	152 (8.6%)	Vitamin deficiency, other	57 (0.3%)	7 (0.4%)
Vocal cord paralysis	271 (1.4%)	20 (1.1%)	Von Willebrand's disease	39 (0.2%)	4 (0.2%)
Vulvodynia	5 (0.0%)	0 (0.0%)	Wheelchair dependence	512 (2.7%)	39 (2.2%)
White blood cell disorder, other	383 (2.0%)	64 (3.6%)	Williams syndrome	11 (0.1%)	0 (0.0%)
Wiskott-Aldrich syndrome	7 (0.0%)	0 (0.0%)	Xeroderma pigmentosum	8 (0.0%)	0 (0.0%)
vesico-uretero-renal reflux	507 (2.6%)	73 (4.1%)			

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	**Primary Procedure, no. %**	Training data (n=19281)	Test data (n=1777)
ABDOMINAL MASS RESECTION, LAPAROSCOPIC, GENSURG	6 (0.5%)	1 (0.2%)	ABLATION WITH CO2 LASER, GENSURG	1 (0.1%)	0 (0.0%)
ABR TESTING, ORL	2 (0.2%)	3 (0.7%)	ACHILLES LENGTHENING, OPEN, ORTHO	52 (4.1%)	6 (1.4%)
ACHILLES LENGTHENING, PERCUTANEOUS, ORTHO	3 (0.2%)	2 (0.5%)	ADENOIDECTOMY REVISION, ORL	2 (0.2%)	0 (0.0%)
ADENOIDECTOMY, ORL	149 (11.7%)	15 (3.4%)	ADJACENT TISSUE ADVANCEMENT CLOSURE, GENSURG	1 (0.1%)	0 (0.0%)
ADJACENT TISSUE TRANSFER, COMPLEX, PLASTICS	11 (0.9%)	0 (0.0%)	ADJACENT TISSUE TRANSFER, SIMPLE, PLASTICS	1 (0.1%)	2 (0.5%)
ADRENALECTOMY, LAPAROSCOPIC, 3MM, GENSURG	1 (0.1%)	0 (0.0%)	ADRENALECTOMY, LAPAROSCOPIC, 5MM, GENSURG	9 (0.7%)	3 (0.7%)
ADRENALECTOMY, RADICAL, GENSURG	9 (0.7%)	2 (0.5%)	ALLOGRAFT NON-UNION, FRACTURE RECONSTRUCTION, ORTHO	12 (0.9%)	2 (0.5%)
AMPUTATION ABOVE KNEE, ORTHO	2 (0.2%)	0 (0.0%)	AMPUTATION TRANSMETATARSAL, ORTHO	1 (0.1%)	0 (0.0%)
ANAL DILATION, GENSURG	1 (0.1%)	0 (0.0%)	ANATOMIC LYMPH NODE DISSECTION, GENSURG	2 (0.2%)	0 (0.0%)
ANKLE EXPLORATION, ORTHO	3 (0.2%)	0 (0.0%)	ANKLE KIDNER, ORTHO	1 (0.1%)	0 (0.0%)
ANKLE, ARTHROSCOPY WITH ARTHROTOMY, ORTHO	3 (0.2%)	0 (0.0%)	ANKLE, ARTHROSCOPY, ORTHO	5 (0.4%)	0 (0.0%)
ANKLE, BROSTROM PROCEDURE, ORTHO	4 (0.3%)	1 (0.2%)	ANKLE, CRISMAN SNOOK, ORTHO	1 (0.1%)	0 (0.0%)
ANKLE/FOOT, CLOSED REDUCTION, PERCUTANEOUS PINNING, ORTHO	1 (0.1%)	0 (0.0%)	ANKLE/FOOT, ORIF, ORTHO	27 (2.1%)	1 (0.2%)
ANKLE/FOOT, TENDON TRANSFER, ORTHO	41 (3.2%)	4 (0.9%)	ANKLE/FOOT/KNEE EXPLORATION, ORTHO	3 (0.2%)	0 (0.0%)
ANKLE/FOOT/KNEE, IRRIGATION AND DEBRIDEMENT/VAC DRESSING, ORTHO	2 (0.2%)	0 (0.0%)	ANKLE/FOOT/TOE TENDON LENGTHENING, ORTHO	3 (0.2%)	2 (0.5%)
ANOPLASTY, PERINEAL, GENSURG	7 (0.5%)	2 (0.5%)	ANOPROCTOSCOPY, GENSURG	1 (0.1%)	0 (0.0%)
ANORECTAL MANOMETRY, GI	5 (0.4%)	0 (0.0%)	ANORECTOPLASTY, GENSURG	69 (5.4%)	17 (3.8%)
ANORECTOPLASTY, POSTERIOR SAGITTAL, LAPAROSCOPIC ASSISTED, 3MM, GENSURG	2 (0.2%)	1 (0.2%)	ANORECTOPLASTY, POSTERIOR SAGITTAL, LAPAROSCOPIC ASSISTED, 5MM, GENSURG	3 (0.2%)	0 (0.0%)
ANOSCOPY WITH RECTAL DILATATION, GENSURG	1 (0.1%)	0 (0.0%)	ANTERIOR AORTOPEXY/TRACHEOPEXY, OPEN, GENSURG	18 (1.4%)	0 (0.0%)
AORTIC BYPASS, GENSURG	2 (0.2%)	0 (0.0%)	AORTOPEXY, OPEN, GENSURG	4 (0.3%)	0 (0.0%)
APERT FOOT RECONSTRUCTION, PLASTICS	3 (0.2%)	0 (0.0%)	APERT HAND RECONSTRUCTION, PLASTICS	6 (0.5%)	0 (0.0%)
APPENDECTOMY, INTERVAL, LAPAROSCOPIC, GENSURG	53 (4.2%)	0 (0.0%)	APPENDECTOMY, LAPAROSCOPIC, 3MM, GENSURG	3 (0.2%)	0 (0.0%)
APPENDECTOMY, LAPAROSCOPIC, 5MM, GENSURG	12 (0.9%)	3 (0.7%)	APPENDICOCECOSTOMY (ACE PROCEDURE), GENSURG	18 (1.4%)	1 (0.2%)
APPENDICOSTOMY REVISION, OPEN, GENSURG	4 (0.3%)	0 (0.0%)	APPENDICOSTOMY, GENSURG	1 (0.1%)	0 (0.0%)
APPENDICOSTOMY, LAPAROSCOPIC, 3MM, GENSURG	2 (0.2%)	0 (0.0%)	APPENDICOSTOMY, LAPAROSCOPIC, 5MM, GENSURG	25 (2.0%)	3 (0.7%)
APPENDICOSTOMY/ACE PROCEDURE, ROBOTIC, GENSURG	1 (0.1%)	1 (0.2%)	APPENDICOVESICOSTOMY, GU	10 (0.8%)	2 (0.5%)
ARTHROGRAM AND/OR INJECTION, ORTHO	3 (0.2%)	0 (0.0%)	ARTHROSCOPY HIP OSTEOPLASTY COMPLEX, ORTHO	14 (1.1%)	0 (0.0%)
ARTHROSCOPY HIP OSTEOPLASTY SIMPLE, ORTHO	14 (1.1%)	0 (0.0%)	ARTHROSCOPY KNEE ACL RECONSTRUCTION, REVISION, ORTHO	1 (0.1%)	0 (0.0%)
ARTHROSCOPY KNEE WITH TRILLAT, ORTHO	39 (3.1%)	0 (0.0%)	ARTIFICIAL URINARY SPHINCTER IMPLANT, GU	1 (0.1%)	1 (0.2%)
AUTOGENOUS BONE GRAFT, PLASTICS	1 (0.1%)	0 (0.0%)	AV FISTULA CREATION/REVISION, GENSURG	21 (1.6%)	0 (0.0%)
AVM/LM EXCISION, ORL	5 (0.4%)	0 (0.0%)	AXILLARY NODE EXCISION, GENSURG	2 (0.2%)	0 (0.0%)
BACLOFEN PUMP INSERTION/REVISION, NEURO	101 (7.9%)	9 (2.0%)	BACLOFEN PUMP REMOVAL, NEURO	5 (0.4%)	0 (0.0%)
BASIC EXCISION, COMPLEX, ORL	1 (0.1%)	0 (0.0%)	BASIC EXCISION, SIMPLE, ORL	2 (0.2%)	0 (0.0%)
BIOPSY LIVER, LAPAROSCOPIC, 5MM, GENSURG	1 (0.1%)	0 (0.0%)	BIOPSY LUNG, LAPAROSCOPIC, GENSURG	1 (0.1%)	0 (0.0%)
BIOPSY LUNG, THORACOSCOPIC, 3MM, GENSURG	1 (0.1%)	1 (0.2%)	BIOPSY LUNG, THORACOSCOPIC, 5MM, GENSURG	40 (3.1%)	6 (1.4%)
BIOPSY LYMPH NODE, GENSURG	3 (0.2%)	0 (0.0%)	BIOPSY MANDIBLE/MAXILLA , ORL	1 (0.1%)	0 (0.0%)
BIOPSY MANDIBLE/MAXILLA, PLASTICS	2 (0.2%)	0 (0.0%)	BIOPSY RECTAL, GENSURG	1 (0.1%)	0 (0.0%)
BIOPSY SENTINEL LYMPH NODE, GENSURG	1 (0.1%)	2 (0.5%)	BIOPSY SOFT TISSUE, GENSURG	1 (0.1%)	0 (0.0%)
BLADDER AUGMENTATION, GU	8 (0.6%)	1 (0.2%)	BLADDER DIVERTICULECTOMY, GU	5 (0.4%)	0 (0.0%)
BLADDER NECK RECONSTRUCTION, GU	10 (0.8%)	1 (0.2%)	BLADDER NECK SUSPENSION WITH FASCIAL SLING, GU	1 (0.1%)	0 (0.0%)
BLOCK, ABDOMINAL WALL/TAP/SHEATH/RECTUS TPI, PAIN	1 (0.1%)	1 (0.2%)	BLOCK, LUMBAR MEDICAL BRANCH/FACET, FLUOROSCOPIC GUIDANCE, PAIN	1 (0.1%)	0 (0.0%)
BONE CYST CURETTAGE/PACKING, ORTHO	3 (0.2%)	0 (0.0%)	BONE GRAFT, ANTERIOR ILIAC CREST, PLASTICS	3 (0.2%)	0 (0.0%)
BONE MARROW ASPIRATE UNILATERAL, ONCOLOGY	1 (0.1%)	0 (0.0%)	BONE MARROW ASPIRATION AND BIOPSY, BILATERAL, ONCOLOGY	3 (0.2%)	0 (0.0%)
BONE MARROW ASPIRATION AND BIOPSY, ONCOLOGY	5 (0.4%)	0 (0.0%)	BONE MARROW HARVEST, ONCOLOGY	18 (1.4%)	0 (0.0%)
BONE RADICAL RESECTION, ORTHO	3 (0.2%)	0 (0.0%)	BONE RADICAL RESECTION, SIMPLE, ORTHO	3 (0.2%)	0 (0.0%)
BONE, BIOPSY, ORTHO	3 (0.2%)	1 (0.2%)	BONE, BIOPSY/CURETTAGE, COMPLEX, ORTHO	5 (0.4%)	1 (0.2%)
BONE, BIOPSY/CURETTAGE, SIMPLE, ORTHO	3 (0.2%)	0 (0.0%)	BONE, BIOPSY/CURETTAGE/PACKING, COMPLEX, ORTHO	15 (1.2%)	1 (0.2%)
BONE, BIOPSY/CURETTAGE/PACKING, SIMPLE, ORTHO	8 (0.6%)	2 (0.5%)	BONE, RADICAL RESECTION AND RECONSTRUCTION W/ALLOGRAFT, ORTHO	23 (1.8%)	0 (0.0%)
BONE, RADICAL RESECTION AND RECONSTRUCTION W/METAL PROSTHESIS, ORTHO	16 (1.3%)	1 (0.2%)	BONE, RADICAL RESECTION WITH SOFT TISSUE RECONSTRUCTION, ORTHO	2 (0.2%)	1 (0.2%)
BONE, RECONSTRUCTION/REVISION W/METAL PROSTHESIS, ORTHO	12 (0.9%)	1 (0.2%)	BOTOX INJECTION, EYE	1 (0.1%)	0 (0.0%)
BOTOX INJECTION, ORL	2 (0.2%)	0 (0.0%)	BOTOX INJECTION, ORTHO	1 (0.1%)	1 (0.2%)
BOTOX/PHENOL INJECTION, ORTHO	24 (1.9%)	0 (0.0%)	BOWEL RESECTION SMALL BOWEL, GENSURG	5 (0.4%)	0 (0.0%)
BOWEL RESECTION, ADULT, OPEN, GENSURG	4 (0.3%)	1 (0.2%)	BOWEL RESECTION, INFANT, OPEN, GENSURG	1 (0.1%)	0 (0.0%)
BOWEL RESECTION, ROBOTIC, GENSURG	2 (0.2%)	0 (0.0%)	BOWEL RESECTION, TRANSANAL, LAPAROSCOPIC, 3MM, GENSURG	1 (0.1%)	0 (0.0%)
BOWEL RESECTION, TRANSANAL, LAPAROSCOPIC, 5MM, GENSURG	5 (0.4%)	0 (0.0%)	BRACHIAL PLEXUS, WITH SURAL NERVE GRAFT, ORTHO	26 (2.0%)	1 (0.2%)
BRACHIAL PLEXUS, WITHOUT SURAL NERVE GRAFT, ORTHO	5 (0.4%)	1 (0.2%)	BRANCHIAL CLEFT CYST EXCISION, ORL	22 (1.7%)	0 (0.0%)
BREAST AUGMENTATION, BILATERAL, PLASTICS	9 (0.7%)	0 (0.0%)	BREAST AUGMENTATION, UNILATERAL, PLASTICS	6 (0.5%)	3 (0.7%)
BREAST MASS EXCISION, GENSURG	2 (0.2%)	0 (0.0%)	BREAST REDUCTION UNILATERAL, PLASTICS	7 (0.5%)	0 (0.0%)
BREAST REDUCTION WITH LIPOSUCTION, PLASTICS	3 (0.2%)	0 (0.0%)	BREAST REDUCTION, PLASTICS	735 (57.7%)	53 (12.0%)
BRONCHOSCOPY, DYNAMIC, PULMONARY	3 (0.2%)	0 (0.0%)	BRONCHOSCOPY, RIGID, GENSURG	36 (2.8%)	2 (0.5%)
BUR HOLE, NEURO	4 (0.3%)	0 (0.0%)	BUR HOLE/RM 25, NEURO	1 (0.1%)	0 (0.0%)
BURIED PENIS CORRECTION, WITHOUT SKIN GRAFT, PLASTICS	2 (0.2%)	0 (0.0%)	BURR HOLE BIOPSY, NAVIGATION GUIDED, NEURO	3 (0.2%)	1 (0.2%)
CANTHOPLASTY/TRANSNASAL WIRING, PLASTICS	1 (0.1%)	0 (0.0%)	CAPSULOTOMY REMOVAL OR REVISION OF BREAST IMPLANT OR EXPANDER, PLASTICS	2 (0.2%)	0 (0.0%)
CAPSULOTOMY, ORTHO	1 (0.1%)	0 (0.0%)	CARTILAGE GRAFT, RIB, PLASTICS	1 (0.1%)	0 (0.0%)
CASE CANCELED IN OR	4 (0.3%)	0 (0.0%)	CAST APPLICATION/CHANGE, ORTHO	1 (0.1%)	1 (0.2%)
CATARACT/LENS EXTRACTION, BILATERAL, EYE	7 (0.5%)	0 (0.0%)	CATARACT/LENS EXTRACTION, UNILATERAL, EYE	10 (0.8%)	0 (0.0%)
CATARACT/LENS EXTRACTION/BIOMETRY, EYE	12 (0.9%)	1 (0.2%)	CATARACT/LENS EXTRACTION/IOL/ANTERIOR VITRECTOMY/BIOMETRY, EYE	2 (0.2%)	0 (0.0%)
CATARACT/LENS EXTRACTION/IOL/ANTERIOR VITRECTOMY/NO BIOMETRY, EYE	1 (0.1%)	0 (0.0%)	CATARACT/LENS EXTRACTION/IOL/BIOMETRY, EYE	3 (0.2%)	0 (0.0%)
CATHETER PLACEMENT, FEMORAL NERVE SHEATH, PAIN	5 (0.4%)	0 (0.0%)	CATHETER PLACEMENT, INTERSCALENE BRACHIAL PLEXUS, PAIN	2 (0.2%)	0 (0.0%)
CATHETER PLACEMENT, PARAVERTEBRAL, PAIN	1 (0.1%)	0 (0.0%)	CATHETER PLACEMENT, PERIPHERAL NERVE, PAIN	1 (0.1%)	1 (0.2%)
CATHETER PLACEMENT, POSTERIOR LUMBAR PLEXUS/FLUOROSCOPIC GUIDANCE, PAIN	1 (0.1%)	0 (0.0%)	CATHETER PLACEMENT, SCIATIC NERVE, PAIN	5 (0.4%)	1 (0.2%)
CECOSTOMY TUBE CHANGE, GI	1 (0.1%)	0 (0.0%)	CECOSTOMY TUBE PLACEMENT, LAPAROSCOPIC, 3MM, GENSURG	3 (0.2%)	1 (0.2%)
CECOSTOMY TUBE PLACEMENT, LAPAROSCOPIC, 5MM, GENSURG	30 (2.4%)	1 (0.2%)	CENTRAL VENOUS LINE INSERTION/REVISION, TUNNELED, GENSURG	58 (4.6%)	2 (0.5%)
CENTRAL VENOUS LINE REVISION, GENSURG	3 (0.2%)	0 (0.0%)	CENTRAL VENOUS LINE/PHERESIS CATHETER/HEMODIALYSIS CATHETER REMOVAL, GENSURG	2 (0.2%)	3 (0.7%)
CHEMOTHERAPY RESERVOIR, OMMAYA, NEURO	11 (0.9%)	1 (0.2%)	CHEST RECONSTRUCTION WITH NIPPLE GRAFTING, PLASTICS	171 (13.4%)	13 (2.9%)
CHEST RECONSTRUCTION WITHOUT NIPPLE GRAFTING, PLASTICS	5 (0.4%)	0 (0.0%)	CHEST WALL RECONSTRUCTION BY MODIFIED RAVITCH PROCEDURE, GENSURG	2 (0.2%)	0 (0.0%)
CHEST WALL RESECTION, GENSURG	3 (0.2%)	0 (0.0%)	CHOANAL ATRESIA REPAIR, ENDOSCOPIC WITH MICRODEBRIDER, ORL	9 (0.7%)	1 (0.2%)
CHOLANGIOGRAM, GENSURG	1 (0.1%)	0 (0.0%)	CHOLECYSTECTOMY (WITH OR WITHOUT CHOLANGIOGRAM), LAPAROSCOPIC, 5MM, GENSURG	112 (8.8%)	14 (3.2%)
CHOLECYSTECTOMY (WITH OR WITHOUT CHOLANGIOGRAM), OPEN, GENSURG	5 (0.4%)	0 (0.0%)	CHOLECYSTECTOMY WITH CHOLANGIOGRAM, LAPAROSCOPIC, GENSURG	3 (0.2%)	0 (0.0%)
CHOLEDOCHAL CYST EXCISION, LAPAROSCOPIC, 5MM, GENSURG	3 (0.2%)	0 (0.0%)	CHOLEDOCHAL CYST EXCISION, OPEN, GENSURG	7 (0.5%)	2 (0.5%)
CHORDEE RELEASE MILD, GU	2 (0.2%)	0 (0.0%)	CHORDEE RELEASE MODERATE, GU	8 (0.6%)	0 (0.0%)
CHORDEE RELEASE SEVERE, GU	2 (0.2%)	0 (0.0%)	CIRCUMCISION REVISION, GU	3 (0.2%)	0 (0.0%)
CIRCUMCISION, GENSURG	5 (0.4%)	1 (0.2%)	CIRCUMCISION, GU	13 (1.0%)	0 (0.0%)
CLAVICLE, ORIF, ORTHO	7 (0.5%)	1 (0.2%)	CLEFT LIP ADHESION/LIP NASAL ADHESION, PLASTICS	25 (2.0%)	0 (0.0%)
CLEFT LIP REPAIR, COMPLETE, BILATERAL, PLASTICS	18 (1.4%)	1 (0.2%)	CLEFT LIP REPAIR, COMPLETE, UNILATERAL, PLASTICS	39 (3.1%)	4 (0.9%)
CLEFT LIP REPAIR, INCOMPLETE, BILATERAL, PLASTICS	4 (0.3%)	2 (0.5%)	CLEFT LIP REPAIR, INCOMPLETE, UNILATERAL, PLASTICS	42 (3.3%)	4 (0.9%)
CLEFT LIP REVISION, PLASTICS	11 (0.9%)	1 (0.2%)	CLEFT NASAL DEFORMITY REVISION, PLASTICS	35 (2.7%)	0 (0.0%)
CLEFT NASAL REVISION, PLASTICS	10 (0.8%)	0 (0.0%)	CLEFT PALATE REPAIR, VEAU I, SUBMUCOUS, PLASTICS	41 (3.2%)	6 (1.4%)
CLEFT PALATE REPAIR, VEAU II, PLASTICS	9 (0.7%)	3 (0.7%)	CLEFT PALATE REPAIR, VEAU III, UNILATERAL, PLASTICS	56 (4.4%)	8 (1.8%)
CLEFT PALATE REPAIR, VEAU IV, BILATERAL, PLASTICS	45 (3.5%)	3 (0.7%)	CLEFT PALATE REVISION, PLASTICS	7 (0.5%)	0 (0.0%)
CLEFT SOFT PALATE REPAIR, PLASTICS	5 (0.4%)	0 (0.0%)	CLITOROPLASTY, GU	7 (0.5%)	0 (0.0%)
CLOACAL RECONSTRUCTION, 9 KG TO 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	CLOACAL RECONSTRUCTION, ¡ 9 KG, GENSURG	2 (0.2%)	0 (0.0%)
CLOACAL RECONSTRUCTION, ¿ 25 KG , GENSURG	1 (0.1%)	0 (0.0%)	CLOACAL RECONSTRUCTION, GU	4 (0.3%)	0 (0.0%)
CLOSED REDUCTION HIP WITH ARTHROGRAM AND SPICA CAST/MRT, ORTHO	2 (0.2%)	0 (0.0%)	CLUB FOOT CORRECTION, ORTHO	2 (0.2%)	0 (0.0%)
COCCYGECTOMY, ORTHO	1 (0.1%)	0 (0.0%)	COCHLEAR IMPLANT BILATERAL BASIC, ORL	20 (1.6%)	6 (1.4%)
COCHLEAR IMPLANT BILATERAL COMPLEX, ORL	31 (2.4%)	1 (0.2%)	COCHLEAR IMPLANT REMOVAL, ORL	4 (0.3%)	1 (0.2%)
COCHLEAR IMPLANT UNILATERAL BASIC, ORL	68 (5.3%)	6 (1.4%)	COCHLEAR IMPLANT UNILATERAL COMPLEX, ORL	57 (4.5%)	3 (0.7%)
COCHLEAR REIMPLANT REVISION, ORL	4 (0.3%)	0 (0.0%)	COLECTOMY (WITH OR WITHOUT ILEOSTOMY), LAPAROSCOPIC ASSISTED, 3MM, GENSURG	1 (0.1%)	0 (0.0%)
COLECTOMY (WITH OR WITHOUT ILEOSTOMY), LAPAROSCOPIC ASSISTED, 5MM, GENSURG	20 (1.6%)	0 (0.0%)	COLECTOMY (WITH OR WITHOUT ILEOSTOMY), LAPAROSCOPIC ASSISTED, ¿ 25 KG, GENSURG	2 (0.2%)	0 (0.0%)
COLECTOMY, LAPAROSCOPIC, GENSURG	5 (0.4%)	0 (0.0%)	COLECTOMY, PARTIAL, 9 TO 25 KG, GENSURG	4 (0.3%)	0 (0.0%)
COLECTOMY, PARTIAL, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	COLECTOMY, ROBOTIC, GENSURG	1 (0.1%)	0 (0.0%)
COLECTOMY, SUBTOTAL, LAPAROSCOPIC, 5MM, GENSURG	10 (0.8%)	0 (0.0%)	COLECTOMY, TOTAL ABDOMINAL and ILEOSTOMY, 9 TO 25 KG, GENSURG	1 (0.1%)	1 (0.2%)
COLECTOMY, TOTAL ABDOMINAL and ILEOSTOMY, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	COLECTOMY, TRANSANAL, LAPAROSCOPIC ASSIST, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)
COLECTOMY, TRANSANAL, LAPAROSCOPIC ASSISTED, 5MM, GENSURG	4 (0.3%)	0 (0.0%)	COLONOSCOPY WITH CATHETER PLACEMENT AND FLUOROSCOPY, GI	157 (12.3%)	15 (3.4%)
COLONOSCOPY WITH CONTROL OF HEMORRHAGE, GI	1 (0.1%)	0 (0.0%)	COLONOSCOPY WITH POLYPECTOMY, GI	6 (0.5%)	2 (0.5%)
COLONOSCOPY, GI	146 (11.5%)	16 (3.6%)	COLOSTOMY CLOSURE, GENSURG	31 (2.4%)	0 (0.0%)

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)
COLOSTOMY CLOSURE, ROBOTIC, GENSURG	1 (0.1%)	0 (0.0%)	COLOSTOMY REVISION, GENSURG	2 (0.2%)	0 (0.0%)
COLOSTOMY, GENSURG	2 (0.2%)	0 (0.0%)	COLOSTOMY, LAPAROSCOPIC ASSISTED, GENSURG	7 (0.5%)	1 (0.2%)
COLUMELLAR LENGTHENING, PLASTICS	1 (0.1%)	0 (0.0%)	CONSTRICTION RING RELEASE, PLASTICS	1 (0.1%)	0 (0.0%)
CONTINENT URINARY DIVERSION, APPENDIX/MITROFANOFF, GU	17 (1.3%)	0 (0.0%)	CORE DECOMPRESSION, ORTHO	10 (0.8%)	1 (0.2%)
CORNEA, CORNEAL COLLAGEN CROSSLINKING, EYE	2 (0.2%)	0 (0.0%)	CORONOIDECTOMY, INTRAORAL, PLASTICS	4 (0.3%)	0 (0.0%)
CORRECTION OF CONCEALED PENIS, GU	4 (0.3%)	0 (0.0%)	CORRECTION OF PENILE TORSION, GU	1 (0.1%)	0 (0.0%)
CRANIAL BONE GRAFT, PLASTICS	1 (0.1%)	0 (0.0%)	CRANIOPLASTY WITH BONE GRAFT, NEURO	2 (0.2%)	0 (0.0%)
CRANIOPLASTY, NEURO	6 (0.5%)	0 (0.0%)	CRANIOPLASTY, PLASTICS	16 (1.3%)	1 (0.2%)
CRANIOTOMY, ANEURYSM/AVM COMPLEX/RM 25, NEURO	5 (0.4%)	0 (0.0%)	CRANIOTOMY, ANEURYSM/AVM SIMPLE, NEURO	1 (0.1%)	0 (0.0%)
CRANIOTOMY, ANEURYSM/AVM SIMPLE/MRT, NEURO	1 (0.1%)	0 (0.0%)	CRANIOTOMY, ANEURYSM/AVM, NEURO	24 (1.9%)	0 (0.0%)
CRANIOTOMY, BIOPSY/RM 25, NEURO	4 (0.3%)	0 (0.0%)	CRANIOTOMY, CAVERNOUS MALFORMATION RESECTION, SHORT, NEURO	5 (0.4%)	0 (0.0%)
CRANIOTOMY, CAVERNOUS MALFORMATION RESECTION, STANDARD, NEURO	21 (1.6%)	0 (0.0%)	CRANIOTOMY, CHIARI DECOMPRESSION, NEURO	131 (10.3%)	15 (3.4%)
CRANIOTOMY, CORPUS CALLOSTOMY, NEURO	20 (1.6%)	2 (0.5%)	CRANIOTOMY, CORPUS COLLOSTOMY/RM 25, NEURO	4 (0.3%)	0 (0.0%)
CRANIOTOMY, ENCEPHALOCELE REPAIR, SHORT, NEURO	3 (0.2%)	0 (0.0%)	CRANIOTOMY, ENCEPHALOCELE REPAIR, STANDARD, NEURO	3 (0.2%)	0 (0.0%)
CRANIOTOMY, FENESTRATION OF CYST, NEURO	12 (0.9%)	1 (0.2%)	CRANIOTOMY, HEMISPHERECTOMY FUNCTIONAL OR ANATOMICAL, NEURO	9 (0.7%)	0 (0.0%)
CRANIOTOMY, MOYA-MOYA UNILATERAL/RM 25, NEURO	5 (0.4%)	0 (0.0%)	CRANIOTOMY, MOYA-MOYA, PIAL SYNANGIOSIS, BILATERAL, NEURO	45 (3.5%)	0 (0.0%)
CRANIOTOMY, MOYA-MOYA, PIAL SYNANGIOSIS, UNILATERAL, NEURO	28 (2.2%)	0 (0.0%)	CRANIOTOMY, NEURO	16 (1.3%)	0 (0.0%)
CRANIOTOMY, OCCIPITAL, TUMOR/MRT, NEURO	2 (0.2%)	0 (0.0%)	CRANIOTOMY, POSTERIOR FOSSA/MRT, NEURO	9 (0.7%)	0 (0.0%)
CRANIOTOMY, POSTERIOR FOSSA/RM 25, NEURO	1 (0.1%)	0 (0.0%)	CRANIOTOMY, RESECTION OF SEIZURE FOCUS/MRT, NEURO	12 (0.9%)	0 (0.0%)
CRANIOTOMY, SEIZURE FOCUS RESECTION, NEURO	46 (3.6%)	3 (0.7%)	CRANIOTOMY, SUBOCCIPITAL, CHIARI DECOMPRESSION/RM 25, NEURO	14 (1.1%)	0 (0.0%)
CRANIOTOMY, SUBOCCIPITAL, TUMOR COMPLEX, NEURO	1 (0.1%)	0 (0.0%)	CRANIOTOMY, TRAUMA, NEURO	1 (0.1%)	0 (0.0%)
CRANIOTOMY, TUMOR BIOPSY, NEURO	26 (2.0%)	0 (0.0%)	CRANIOTOMY, TUMOR COMPLEX, NEURO	2 (0.2%)	0 (0.0%)
CRANIOTOMY, TUMOR, SHORT DURATION, NEURO	37 (2.9%)	7 (1.6%)	CRANIOTOMY, TUMOR, SKULL BASE, NEURO	1 (0.1%)	0 (0.0%)
CRANIOTOMY, TUMOR, STANDARD DURATION, NEURO	93 (7.3%)	12 (2.7%)	CYSTECTOMY PARTIAL, GU	2 (0.2%)	0 (0.0%)
CYSTOLITHOTOMY, GU	3 (0.2%)	0 (0.0%)	CYSTOSCOPY WITH HOLMIUM LASER, GU	1 (0.1%)	0 (0.0%)
CYSTOSCOPY, GENSURG	1 (0.1%)	1 (0.2%)	CYSTOSCOPY, GU	25 (2.0%)	2 (0.5%)
CYSTOSCOPY/BIOPSY, GU	6 (0.5%)	0 (0.0%)	CYSTOSCOPY/CYSTOURETHROGRAPHY, GU	1 (0.1%)	0 (0.0%)
CYSTOSCOPY/DIRECT VISUAL INTERNAL URETHROTOMY (DVIU), GU	4 (0.3%)	0 (0.0%)	CYSTOSCOPY/INCISION OF URETEROCELE, GU	11 (0.9%)	0 (0.0%)
CYSTOSCOPY/LITHIOPAXY WITH HOLMIUM LASER, GU	1 (0.1%)	0 (0.0%)	CYSTOSCOPY/LITHIOPAXY WITH THULIUM LASER, GU	1 (0.1%)	1 (0.2%)
CYSTOSCOPY/RETROGRADE PYELOGRAM, GU	3 (0.2%)	0 (0.0%)	CYSTOSCOPY/SUPRAPUBIC TUBE INSERTION, GU	5 (0.4%)	0 (0.0%)
CYSTOSCOPY/TRANSURETHRAL RESECTION OF BLADDER TUMOR, GU	1 (0.1%)	0 (0.0%)	CYSTOSCOPY/TUR OF VALVES, GU	17 (1.3%)	5 (1.1%)
CYSTOSCOPY/URETERAL CATHETERIZATION, GU	2 (0.2%)	0 (0.0%)	CYSTOSCOPY/URETERAL STENT PLACEMENT, UNILATERAL, GU	1 (0.1%)	1 (0.2%)
CYSTOSCOPY/URETERAL STENT REMOVAL, BILATERAL, GU	2 (0.2%)	0 (0.0%)	CYSTOSCOPY/URETERAL STENT REMOVAL, UNILATERAL, GU	19 (1.5%)	6 (1.4%)
CYSTOSCOPY/URETEROSCOPY, GU	3 (0.2%)	0 (0.0%)	CYSTOSCOPY/URETEROSCOPY/LITHOTRIPSY WITH HOLMIUM LASER, GU	26 (2.0%)	0 (0.0%)
CYSTOSCOPY/URETEROSCOPY/LITHOTRIPSY WITH THULIUM LASER, GU	1 (0.1%)	1 (0.2%)	CYSTOSCOPY/URETEROSCOPY/REMOVAL OF URETERAL STONES, GU	3 (0.2%)	0 (0.0%)
CYSTOSCOPY/URETEROURETEROSTOMY, GU	6 (0.5%)	1 (0.2%)	CYSTOSCOPY/URETHRAL DILATATION, GU	2 (0.2%)	0 (0.0%)
DEBULKING OF FOOT, PLASTICS	1 (0.1%)	0 (0.0%)	DEBULKING OF HAND, COMPLEX, PLASTICS	1 (0.1%)	0 (0.0%)
DEBULKING OF HAND, SIMPLE, PLASTICS	7 (0.5%)	0 (0.0%)	DEBULKING OF HEAD/NECK, COMPLEX, PLASTICS	5 (0.4%)	0 (0.0%)
DEBULKING OF HEAD/NECK, SIMPLE, PLASTICS	1 (0.1%)	0 (0.0%)	DEBULKING OF LOWER EXTREMITY, SIMPLE, PLASTICS	1 (0.1%)	0 (0.0%)
DEBULKING OF TRUNK, COMPLEX, PLASTICS	1 (0.1%)	0 (0.0%)	DEBULKING OF UPPER EXTREMITY, COMPLEX, PLASTICS	6 (0.5%)	0 (0.0%)
DEEP BRAIN STIMULATOR BRAIN LEAD IMPLANT/REVISION, NEURO	6 (0.5%)	1 (0.2%)	DEEP BRAIN STIMULATOR GENERATOR IMPLANT/REVISION, NEURO	4 (0.3%)	2 (0.5%)
DENTAL EXAM, INTRAORAL XRAYS AND SCALING, DENTAL	1 (0.1%)	0 (0.0%)	DENTAL EXTRACTIONS 10-30 MINS, PLASTICS	2 (0.2%)	0 (0.0%)
DENTAL EXTRACTIONS ¡10 MINS, PLASTICS	1 (0.1%)	0 (0.0%)	DENTAL EXTRACTIONS ¿30 MINS, PLASTICS	1 (0.1%)	0 (0.0%)
DENTAL IMPLANTS ¿1 HOUR, PLASTICS	1 (0.1%)	0 (0.0%)	DENTAL MAXILLARY APPLIANCE (DMA) INSERTION, DENTAL	30 (2.4%)	1 (0.2%)
DENTAL MAXILLARY APPLIANCE (DMA) REMOVAL AND IMPRESSION, DENTAL	3 (0.2%)	0 (0.0%)	DENTAL REHAB, ADULT, COMPLEX, DENTAL	4 (0.3%)	0 (0.0%)
DENTAL REHAB, ADULT, SIMPLE, DENTAL	6 (0.5%)	0 (0.0%)	DENTAL REHABILITATION, LONG (¿181 min), DENTAL	3 (0.2%)	0 (0.0%)
DENTAL REHABILITATION, MEDIUM (121 to 180 min), DENTAL	2 (0.2%)	4 (0.9%)	DENTAL REHABILITATION, SHORT (61 to 120 min), DENTAL	21 (1.6%)	4 (0.9%)
DENTAL REHABILITATION, ULTRA SHORT (¡60 min), DENTAL	1 (0.1%)	1 (0.2%)	DERMAL-FAT GRAFT, PLASTICS	8 (0.6%)	0 (0.0%)
DERMOID CYST/SKULL LESION EXCISION, NEURO	10 (0.8%)	0 (0.0%)	DIAPHRAGM PLICATION, LAPAROSCOPIC/THORACOSCOPIC, 3MM, GENSURG	1 (0.1%)	0 (0.0%)
DIAPHRAGMATIC HERNIA REPAIR, MORGAGNI, GENSURG	1 (0.1%)	0 (0.0%)	DIAPHRAGMATIC HERNIA REPAIR, MORGAGNI, THORACOSCOPIC, 3MM, GENSURG	3 (0.2%)	0 (0.0%)
DIAPHRAGMATIC HERNIA REPAIR, MORGAGNI, THORACOSCOPIC, 5MM, GENSURG	2 (0.2%)	0 (0.0%)	DIAPHRAGMATIC HERNIA REPAIR, OPEN, GENSURG	3 (0.2%)	1 (0.2%)
DIAPHRAGMATIC HERNIA REPAIR, THORACOSCOPIC, 5MM, GENSURG	6 (0.5%)	0 (0.0%)	DIODE LASER PHOTOCOAGULATION, EYE	4 (0.3%)	0 (0.0%)
DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY WITH DILATION, ORL	29 (2.3%)	9 (2.0%)	DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY WITH LASER, ORL	8 (0.6%)	1 (0.2%)
DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY WITH MICRODEBRIDER, ORL	14 (1.1%)	0 (0.0%)	DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, INJECTION LARYNGOPLASTY WITH FAT GRAFT, ORL	1 (0.1%)	0 (0.0%)
DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, INJECTION LARYNGOPLASTY, ORL	5 (0.4%)	1 (0.2%)	DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, ORL	685 (53.8%)	72 (16.3%)
DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, SUPRAGLOTTOPLASTY WITH LASER, ORL	5 (0.4%)	1 (0.2%)	DIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, SUPRAGLOTTOPLASTY, ORL	53 (4.2%)	10 (2.3%)
DISTRACTION, MANDIBLE, INFANT, PLASTICS	4 (0.3%)	1 (0.2%)	DISTRACTOR REMOVAL, MANDIBULAR DEVICE, PLASTICS	10 (0.8%)	0 (0.0%)
DISTRACTOR REMOVAL, REDD, PLASTICS	2 (0.2%)	0 (0.0%)	DORSAL RHIZOTOMY, NEURO	9 (0.7%)	0 (0.0%)
DORSAL RHIZOTOMY/RM 25, NEURO	6 (0.5%)	0 (0.0%)	DUODENAL ATRESIA/STRICTURE REPAIR, OPEN, 9 TO 25 KG, GENSURG	1 (0.1%)	0 (0.0%)
DUODENAL ATRESIA/STRICTURE REPAIR, OPEN, ¿ 25 KG, GENSURG	2 (0.2%)	0 (0.0%)	EAR CANAL ATRESIA REPAIR, ORL	3 (0.2%)	0 (0.0%)
EAR RECONSTRUCTION, PLASTICS	12 (0.9%)	0 (0.0%)	ECHO, CARDIAC	1 (0.1%)	0 (0.0%)
ELASTIC CHAIN MAXILLARY RETRACTION (ECPR) APPLIANCE INSERTION, DENTAL	11 (0.9%)	0 (0.0%)	ELBOW, ARTHROSCOPY AND ARTHROTOMY, OATS PROCEDURE, ORTHO	54 (4.2%)	3 (0.7%)
ELBOW, ARTHROSCOPY WITH ARTHROTOMY, ORTHO	9 (0.7%)	0 (0.0%)	ELBOW, ARTHROSCOPY, ORTHO	1 (0.1%)	0 (0.0%)
ELBOW, ARTHROSCOPY, ULNAR COLLATERAL LIGAMENT RECONSTRUCTION, ORTHO	3 (0.2%)	0 (0.0%)	ELBOW, CHRONIC MONTEGGIA, ORTHO	5 (0.4%)	1 (0.2%)
ELBOW, CLOSED REDUCTION, PERCUTANEOUS PINNING, ORTHO	1 (0.1%)	0 (0.0%)	ELBOW, CONTRACTURE RELEASE, ORTHO	6 (0.5%)	0 (0.0%)
ELBOW, EXCISION, RADIAL HEAD, ORTHO	4 (0.3%)	0 (0.0%)	ELBOW, EXPLORATION, OPEN, ORTHO	17 (1.3%)	2 (0.5%)
ELBOW, NERVE TRANSPOSITION, ORTHO	2 (0.2%)	1 (0.2%)	ELECTRORETINOGRAM, EYE	1 (0.1%)	1 (0.2%)
ENCEPHALOCELE REPAIR, PLASTICS	3 (0.2%)	0 (0.0%)	ENDOBRONCHIAL BIOPSY, PULMONARY	1 (0.1%)	0 (0.0%)
ENDODONTIC TREATMENT AND REHABILITATION, LONG (¿181 min), DENTAL	1 (0.1%)	0 (0.0%)	ENDOSCOPIC RETROGRADE CHOLANGIO PANCREATOGRAPHY WITH FLUOROSCOPY, GI	49 (3.8%)	3 (0.7%)
ENDOSCOPY FOR FENESTRATION OF CYST/MRT, NEURO	1 (0.1%)	0 (0.0%)	ENDOSCOPY FOR THIRD VENTRICULOSTOMY WITH BRAIN BIOPSY/RM 25, NEURO	1 (0.1%)	0 (0.0%)
ENDOSCOPY FOR THIRD VENTRICULOSTOMY/RM 25, NEURO	2 (0.2%)	0 (0.0%)	ENDOSCOPY WITH WIRELESS CAPSULE PLACEMENT, GI	2 (0.2%)	0 (0.0%)
ENTEROSCOPY, GI	4 (0.3%)	0 (0.0%)	ENUCLEATION, EYE	13 (1.0%)	2 (0.5%)
EPIDURAL CATHETER PLACEMENT, TUNNELED, PAIN	2 (0.2%)	0 (0.0%)	EPISPADIAS REPAIR, GU	2 (0.2%)	0 (0.0%)
EPISPADIAS REPAIR, PROXIMAL, GU	3 (0.2%)	0 (0.0%)	ESOPHAGEAL ATRESIA REPAIR, INTERPOSITION, GENSURG	13 (1.0%)	0 (0.0%)
ESOPHAGEAL ATRESIA REPAIR, STAGE 1, GENSURG	1 (0.1%)	0 (0.0%)	ESOPHAGEAL ATRESIA REPAIR, STAGE 3, LAPAROSOPIC ASSISTED NISSEN, GENSURG	3 (0.2%)	0 (0.0%)
ESOPHAGEAL ATRESIA REPAIR, STAGE 3, OPEN NISSEN, GENSURG	9 (0.7%)	1 (0.2%)	ESOPHAGEAL STRICTURE REPAIR, GENSURG	8 (0.6%)	0 (0.0%)
ESOPHAGEAL STRICTURE REPAIR, S/P ESOPHAGEAL ATRESIA REPAIR, GENSURG	5 (0.4%)	0 (0.0%)	ESOPHAGOGASTRODUODENOSCOPY WITH BOTOX, GI	16 (1.3%)	3 (0.7%)
ESOPHAGOGASTRODUODENOSCOPY WITH CATHETER PLACEMENT AND FLUOROSCOPY, GI	35 (2.7%)	1 (0.2%)	ESOPHAGOGASTRODUODENOSCOPY WITH CONTROL OF HEMMORHAGE, GI	2 (0.2%)	0 (0.0%)
ESOPHAGOGASTRODUODENOSCOPY WITH DILATION AND FLUOROSCOPY, GI	157 (12.3%)	7 (1.6%)	ESOPHAGOGASTRODUODENOSCOPY WITH ENDOSCOPIC ULTRASOUND, GI	3 (0.2%)	0 (0.0%)
ESOPHAGOGASTRODUODENOSCOPY WITH ESOPHAGEAL STENT PLACEMENT/REMOVAL, GI	1 (0.1%)	1 (0.2%)	ESOPHAGOGASTRODUODENOSCOPY WITH FOREIGN BODY REMOVAL, GI	1 (0.1%)	0 (0.0%)
ESOPHAGOGASTRODUODENOSCOPY WITH IMPEDANCE PROBE PLACEMENT, GI	33 (2.6%)	1 (0.2%)	ESOPHAGOGASTRODUODENOSCOPY WITH PEG, GI	3 (0.2%)	0 (0.0%)
ESOPHAGOGASTRODUODENOSCOPY WITH PER ORAL ENDOSCOPIC MYOTOMY, GI	3 (0.2%)	0 (0.0%)	ESOPHAGOGASTRODUODENOSCOPY WITH TUBE PLACEMENT/CHANGE, GI	6 (0.5%)	0 (0.0%)
ESOPHAGOGASTRODUODENOSCOPY WITH VARICEAL LIGATION OR SCLEROSIS, GI	1 (0.1%)	0 (0.0%)	ESOPHAGOGASTRODUODENOSCOPY, GI	75 (5.9%)	34 (7.7%)
ESOPHAGOMYOTOMY, ROBOTIC, GENSURG	1 (0.1%)	0 (0.0%)	ESOPHAGOMYOTOMY, THORACOSCOPIC, 5MM, GENSURG	3 (0.2%)	0 (0.0%)
ESOPHAGOSCOPY, ORL	1 (0.1%)	0 (0.0%)	EUA, EXCISION/BIOPSY VAGINAL LESION, GYN	1 (0.1%)	0 (0.0%)
EUSTACHIAN TUBE RECONSTRUCTION, ORL	1 (0.1%)	0 (0.0%)	EXAM UNDER ANESTHESIA NASOPHARYNX, ORL	5 (0.4%)	0 (0.0%)
EXAM UNDER ANESTHESIA WITH INTRAVITREAL CHEMO INJECTION, EYE	2 (0.2%)	0 (0.0%)	EXAM UNDER ANESTHESIA, COMPLEX, EYE	32 (2.5%)	0 (0.0%)
EXAM UNDER ANESTHESIA, EAR, ORL	58 (4.6%)	1 (0.2%)	EXAM UNDER ANESTHESIA, GU	3 (0.2%)	0 (0.0%)
EXAM UNDER ANESTHESIA, ONCOLOGY	1 (0.1%)	0 (0.0%)	EXAM UNDER ANESTHESIA, ORL	1 (0.1%)	0 (0.0%)
EXAM UNDER ANESTHESIA, PAIN	1 (0.1%)	0 (0.0%)	EXAM UNDER ANESTHESIA, RECTAL, GENSURG	29 (2.3%)	3 (0.7%)
EXAM UNDER ANESTHESIA, SIMPLE, EYE	4 (0.3%)	1 (0.2%)	EXAM UNDER ANESTHESIA, VAGINAL, GENSURG	2 (0.2%)	0 (0.0%)
EXCISION CYST OF MANDIBLE/MAXILLA ¡1.5 HOURS, PLASTICS	11 (0.9%)	1 (0.2%)	EXCISION CYST OF MANDIBLE/MAXILLA ¿1.5 HOURS, PLASTICS	3 (0.2%)	2 (0.5%)
EXCISION DERMOID CYST, INTRACRANIAL, PLASTICS	2 (0.2%)	1 (0.2%)	EXCISION LESION MEDIUM, PLASTICS	7 (0.5%)	0 (0.0%)
EXCISION LESION, GU	1 (0.1%)	0 (0.0%)	EXCISION MASS, LARGE, PLASTICS	6 (0.5%)	0 (0.0%)
EXCISION MASS, SMALL, PLASTICS	1 (0.1%)	0 (0.0%)	EXCISION PATENT URACHUS/REMOVAL URACHAL CYST, GU	4 (0.3%)	0 (0.0%)
EXCISION SKIN LESION LARGE, PLASTICS	15 (1.2%)	2 (0.5%)	EXCISION SKIN LESION SMALL, PLASTICS	10 (0.8%)	1 (0.2%)
EXCISION URETEROCELE, GU	1 (0.1%)	2 (0.5%)	EXCISION VASCULAR MALFORMATION (AVM/LM/VM) ¡1.5 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)
EXCISION VASCULAR MALFORMATION (AVM/LM/VM) ¿1.5 HOURS, PLASTICS	12 (0.9%)	0 (0.0%)	EXCISION VASCULAR MALFORMATION (AVM/LM/VM) COMPLICATED ¿4 HOURS, PLASTICS	16 (1.3%)	0 (0.0%)
EXCISION VASCULAR MALFORMATION (AVM/LM/VM), EXTREMITY, LARGE, PLASTICS	1 (0.1%)	1 (0.2%)	EXOSTOSIS EXCISION, ORTHO	8 (0.6%)	0 (0.0%)
EXSTROPHY OF BLADDER CLOSURE, GU	7 (0.5%)	0 (0.0%)	EXTERNAL FIXATION APPLICATION FOR FRACTURE, ORTHO	5 (0.4%)	1 (0.2%)
EXTERNAL VENTRICULAR DRAIN (EVD) PLACEMENT, NEURO	3 (0.2%)	0 (0.0%)	EXTRACORPOREAL SHOCK WAVE LITHOTRIPSY, GU	8 (0.6%)	1 (0.2%)
EYE TRAUMA AND/ OR FOREIGN BODY REPAIR, EYE	1 (0.1%)	0 (0.0%)	EYELID LESION EXCISION, EYE	1 (0.1%)	0 (0.0%)
FACIAL FRACTURE, ORIF ¿2 HOURS, PLASTICS	1 (0.1%)	1 (0.2%)	FASCIOTOMY, ORTHO	1 (0.1%)	0 (0.0%)
FASCIOTOMY, PLASTICS	1 (0.1%)	0 (0.0%)	FAT GRAFTING, LIPOSUCTION, PLASTICS	6 (0.5%)	0 (0.0%)
FECAL DISIMPACTION UNDER ANESTHESIA, GI	4 (0.3%)	0 (0.0%)	FEMORAL ARTERY RECONSTRUCTION, GENSURG	2 (0.2%)	0 (0.0%)
FEMUR DISTAL, OSTEOTOMY, ORTHO	4 (0.3%)	0 (0.0%)	FEMUR PROXIMAL, OSTEOTOMY, BILATERAL, ORTHO	34 (2.7%)	0 (0.0%)

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)
FEMUR PROXIMAL, OSTEOTOMY, UNILATERAL, ORTHO	95 (7.5%)	9 (2.0%)	FEMUR, HARDWARE REMOVAL, BILATERAL, ORTHO	65 (5.1%)	8 (1.8%)
FEMUR, HARDWARE REMOVAL, UNILATERAL, ORTHO	28 (2.2%)	4 (0.9%)	FEMUR, HEAD RESECTION, CASTLE, ORTHO	1 (0.1%)	0 (0.0%)
FEMUR, HEAD RESECTION, MCHALE, ORTHO	1 (0.1%)	0 (0.0%)	FEMUR, IRRIGATION AND DEBRIDEMENT/VAC DRESSING, ORTHO	1 (0.1%)	0 (0.0%)
FEMUR, ORIF WITH IM NAIL/IM ROD, ORTHO	8 (0.6%)	3 (0.7%)	FEMUR, ORIF, ORTHO	9 (0.7%)	1 (0.2%)
FEMUR, OSTEOTOMY WEDGE WITH PUDDU PLATE, ORTHO	2 (0.2%)	0 (0.0%)	FEMUR, OSTEOTOMY WITH EXTERNAL FIXATOR, ORTHO	3 (0.2%)	0 (0.0%)
FEMUR, OSTEOTOMY WITH IM NAIL/IM ROD, ORTHO	24 (1.9%)	1 (0.2%)	FEMUR, PROXIMAL RESECTION, ORTHO	1 (0.1%)	0 (0.0%)
FEMUR, SOFIELD OSTEOTOMY, ORTHO	11 (0.9%)	2 (0.5%)	FEMUR, TRACTION APPLICATION, ORTHO	1 (0.1%)	0 (0.0%)
FENESTRATION OF CYST, ENDOSCOPY, NEURO	18 (1.4%)	2 (0.5%)	FESS COMPLEX, ORL	41 (3.2%)	2 (0.5%)
FESS LIMITED, ORL	18 (1.4%)	2 (0.5%)	FISTULA IN ANO PROCEDURE, GENSURG	1 (0.1%)	0 (0.0%)
FLEXIBLE BRONCHOSCOPY/BRONCHIOALVEOLAR LAVAGE (BAL), PULMONARY	94 (7.4%)	9 (2.0%)	FOOT RECONSTRUCTION POLYDACTYLY, ORTHO	3 (0.2%)	0 (0.0%)
FOOT RECONSTRUCTION SOFT TISSUE, ORTHO	7 (0.5%)	0 (0.0%)	FOOT, AMPUTATION, ORTHO	15 (1.2%)	2 (0.5%)
FOOT, ARTHRODESIS MIDFOOT/SUBTALAR, ORTHO	11 (0.9%)	1 (0.2%)	FOOT, ARTHRODESIS TOES, ORTHO	16 (1.3%)	3 (0.7%)
FOOT, ARTHRODESIS TRIPLE, ANKLE, ORTHO	16 (1.3%)	2 (0.5%)	FOOT, BUNION RECONSTRUCTION, ORTHO	3 (0.2%)	0 (0.0%)
FOOT, CAPSULOTOMY MIDFOOT/POSTEROMEDIAL RELEASE, ORTHO	24 (1.9%)	2 (0.5%)	FOOT, CAVOVARUS RECONSTRUCTION, ORTHO	67 (5.3%)	3 (0.7%)
FOOT, COALITION EXCISION, ORTHO	14 (1.1%)	0 (0.0%)	FOOT, FLATFOOT RECONSTRUCTION, ORTHO	8 (0.6%)	1 (0.2%)
FOOT, FLATFOOT RECONSTRUCTION, SESA SCREW, ORTHO	1 (0.1%)	0 (0.0%)	FOOT, ORIF LISFRANC JOINT, ORTHO	1 (0.1%)	0 (0.0%)
FOOT, OS TRIGONUM EXCISION, ORTHO	2 (0.2%)	0 (0.0%)	FOOT, OSTEOTOMY, FOREFOOT, ORTHO	2 (0.2%)	2 (0.5%)
FOOT, OSTEOTOMY, HINDFOOT, ORTHO	17 (1.3%)	4 (0.9%)	FOOT, OSTEOTOMY, MIDFOOT, ORTHO	7 (0.5%)	1 (0.2%)
FOOT, POLYDACTYLY EXCISION, ORTHO	3 (0.2%)	0 (0.0%)	FOOT, SOFT TISSUE RECONSTRUCTION, COMPLEX, ORTHO	22 (1.7%)	0 (0.0%)
FOOT, SOFT TISSUE RECONSTRUCTION, SIMPLE, ORTHO	3 (0.2%)	0 (0.0%)	FOOT/TOE, TENOTOMY, ORTHO	1 (0.1%)	0 (0.0%)
FOREARM, ARTHRODESIS, OPEN, ORTHO	1 (0.1%)	0 (0.0%)	FOREARM, EXFIX, ORTHO	3 (0.2%)	0 (0.0%)
FOREARM, OSTEOTOMY, ORTHO	66 (5.2%)	2 (0.5%)	FOREARM, TENDON LENGTHENING, ORTHO	3 (0.2%)	0 (0.0%)
FOREARM, TENDON TRANSFER, ORTHO	3 (0.2%)	1 (0.2%)	FOREARM/WRIST, AMPUTATION, ORTHO	1 (0.1%)	0 (0.0%)
FOREIGN BODY REMOVAL, ORTHO	2 (0.2%)	0 (0.0%)	FOREIGN BODY REMOVAL, SOFT TISSUE, GENSURG	1 (0.1%)	0 (0.0%)
FREE FLAP, FACIAL REANIMATION WITH GRACILIS, PLASTICS	2 (0.2%)	0 (0.0%)	FREE FLAP, FIBULA, PLASTICS	1 (0.1%)	0 (0.0%)
FREE FLAP, SOFT TISSUE, PLASTICS	5 (0.4%)	2 (0.5%)	FRENULOTOMY, LINGUAL, ORL	3 (0.2%)	0 (0.0%)
FRENULOTOMY, UPPER LIP, ORL	1 (0.1%)	0 (0.0%)	FRENULOTOMY, UPPER LIP, PLASTICS	1 (0.1%)	0 (0.0%)
FRONTAL ORBITAL ADVANCEMENT, PLASTICS	50 (3.9%)	6 (1.4%)	GASTRECTOMY, SLEEVE, LAPAROSCOPIC, GENSURG	20 (1.6%)	0 (0.0%)
GASTRIC BYPASS, LAPAROSCOPIC, GENSURG	24 (1.9%)	4 (0.9%)	GASTROCUTANEOUS FISTULA CLOSURE, GENSURG	54 (4.2%)	5 (1.1%)
GASTROJEJUNAL (GJ) TUBE CHANGE, GENSURG	1 (0.1%)	0 (0.0%)	GASTROJEJUNAL (GJ) TUBE PLACEMENT, OPEN, 9 TO 25 KG, GENSURG	3 (0.2%)	0 (0.0%)
GASTROJEJUNAL (GJ) TUBE PLACEMENT, OPEN, ¿ 25 KG, GENSURG	2 (0.2%)	0 (0.0%)	GASTROJEJUNOSTOMY, LAPAROSCOPIC, 3MM, GENSURG	1 (0.1%)	0 (0.0%)
GASTROJEJUNOSTOMY, LAPAROSCOPIC, 5MM, GENSURG	4 (0.3%)	1 (0.2%)	GASTROSTOMY (G) TUBE CHANGE, GENSURG	5 (0.4%)	1 (0.2%)
GASTROSTOMY CLOSURE, GENSURG	1 (0.1%)	0 (0.0%)	GASTROSTOMY, LAPAROSCOPIC, 3MM, GENSURG	51 (4.0%)	18 (4.1%)
GASTROSTOMY, LAPAROSCOPIC, 5MM, GENSURG	219 (17.2%)	8 (1.8%)	GASTROSTOMY, OPEN, GENSURG	4 (0.3%)	0 (0.0%)
GINGIVOPLASTY, PLASTICS	1 (0.1%)	0 (0.0%)	GLOSSECTOMY, ORL	5 (0.4%)	0 (0.0%)
GLOSSECTOMY, PLASTICS	1 (0.1%)	1 (0.2%)	GONIOTOMY, EYE	2 (0.2%)	0 (0.0%)
GROWING SPINAL RODS REMOVAL, ORTHO	3 (0.2%)	0 (0.0%)	GROWING SPINAL RODS REVISION, ORTHO	15 (1.2%)	0 (0.0%)
GYNECOMASTIA REPAIR, MASTECTOMY, GENSURG	1 (0.1%)	0 (0.0%)	GYNECOMASTIA REPAIR, MASTECTOMY, LIPOSUCTION, PLASTICS	14 (1.1%)	1 (0.2%)
GYNECOMASTIA REPAIR, MASTECTOMY, PLASTICS	4 (0.3%)	0 (0.0%)	GYNECOMASTIA REPAIR, SIMPLE MASTECTOMY, UNILATERAL, PLASTICS	3 (0.2%)	0 (0.0%)
HAND EXPLORATION, ORTHO	2 (0.2%)	0 (0.0%)	HAND RECONSTRUCTION SYNDACTYLY SIMPLE, ORTHO	9 (0.7%)	0 (0.0%)
HAND RECONSTRUCTION, COMPLEX, PLASTICS	2 (0.2%)	0 (0.0%)	HAND, CARPAL TUNNEL RELEASE, ORTHO	3 (0.2%)	0 (0.0%)
HAND, CLOSED REDUCTION, PERCUTANEOUS PINNING, ORTHO	1 (0.1%)	0 (0.0%)	HAND, CONGENITAL POLLICIZATION, ORTHO	17 (1.3%)	1 (0.2%)
HAND, CONGENITAL RECONSTRUCTION, LESS THAN 120 MINUTES, ORTHO	40 (3.1%)	1 (0.2%)	HAND, CONGENITAL RECONSTRUCTION, MORE THAN 120 MINUTES, ORTHO	48 (3.8%)	4 (0.9%)
HAND, EPIPHYSIODESIS, ORTHO	1 (0.1%)	0 (0.0%)	HAND, FUSION, ORTHO	1 (0.1%)	0 (0.0%)
HAND, LIGAMENT REPAIR/RECONSTRUCTION, ORTHO	1 (0.1%)	0 (0.0%)	HAND, NERVE REPAIR, ORTHO	1 (0.1%)	0 (0.0%)
HAND, ORIF, ORTHO	1 (0.1%)	0 (0.0%)	HAND, OSTEOTOMY, MULTIPLE, ORTHO	1 (0.1%)	0 (0.0%)
HAND, OSTEOTOMY, SINGLE, ORTHO	3 (0.2%)	0 (0.0%)	HAND, RECONSTRUCTION, ORTHO	48 (3.8%)	0 (0.0%)
HAND, TENDON LENGTHENING, ORTHO	5 (0.4%)	0 (0.0%)	HAND, TENDON REPAIR, ORTHO	1 (0.1%)	0 (0.0%)
HAND, TENDON TRANSFER, ORTHO	5 (0.4%)	0 (0.0%)	HAND, TRAUMA, ORTHO	3 (0.2%)	0 (0.0%)
HAND, Z-PLASTY, ORTHO	1 (0.1%)	0 (0.0%)	HAND/FINGER, AMPUTATION, ORTHO	1 (0.1%)	0 (0.0%)
HAND/FOOT DISTRACTION APPARATUS REMOVAL, PLASTICS	1 (0.1%)	0 (0.0%)	HARDWARE REMOVAL BURIED PIN, PLATE OR SCREW, ORTHO	26 (2.0%)	0 (0.0%)
HARDWARE REMOVAL EXTERNAL FIXATION, ORTHO	3 (0.2%)	0 (0.0%)	HARDWARE REMOVAL FEMUR DISTAL COMPLEX, ORTHO	3 (0.2%)	0 (0.0%)
HARDWARE REMOVAL FEMUR PROXIMAL BILATERAL SIMPLE, ORTHO	8 (0.6%)	0 (0.0%)	HARDWARE REMOVAL FEMUR PROXIMAL UNILATERAL COMPLEX, ORTHO	8 (0.6%)	0 (0.0%)
HARDWARE REMOVAL FEMUR PROXIMAL UNILATERAL SIMPLE, ORTHO	7 (0.5%)	0 (0.0%)	HARDWARE REMOVAL HAND, ORTHO	1 (0.1%)	0 (0.0%)
HARDWARE REMOVAL HAND, PLASTICS	1 (0.1%)	0 (0.0%)	HARDWARE REMOVAL ILIAC SCREW COMPLEX, ORTHO	2 (0.2%)	0 (0.0%)
HARDWARE REMOVAL IM ROD, ORTHO	3 (0.2%)	0 (0.0%)	HARDWARE REMOVAL MAXILLOFACIAL, PLASTICS	9 (0.7%)	2 (0.5%)
HARDWARE REMOVAL TIBIA/FIBULA, ORTHO	3 (0.2%)	0 (0.0%)	HEMIPELVECTOMY, ORTHO	1 (0.1%)	0 (0.0%)
HEMISPHERECTOMY FUNCTIONAL/MRT, NEURO	5 (0.4%)	0 (0.0%)	HEPATIC RESECTION, GENSURG	24 (1.9%)	4 (0.9%)
HERNIA REPAIR, EPIGASTRIC, GENSURG	1 (0.1%)	0 (0.0%)	HERNIA REPAIR, INGUINAL (WITH MESH), LAPAROSCOPIC, GENSURG	1 (0.1%)	0 (0.0%)
HERNIA REPAIR, INGUINAL, BILATERAL, GU	4 (0.3%)	0 (0.0%)	HERNIA REPAIR, INGUINAL, BILATERAL, LAPAROSCOPIC, GENSURG	9 (0.7%)	0 (0.0%)
HERNIA REPAIR, INGUINAL, LAPAROSCOPIC, 3MM, GENSURG	13 (1.0%)	4 (0.9%)	HERNIA REPAIR, INGUINAL, LAPAROSCOPIC, 5MM, GENSURG	6 (0.5%)	0 (0.0%)
HERNIA REPAIR, INGUINAL, UNILATERAL, GU	17 (1.3%)	5 (1.1%)	HERNIA REPAIR, INGUINAL, WITH LAPAROSCOPIC GROIN EXPLORATION, 3MM, GENSURG	9 (0.7%)	3 (0.7%)
HERNIA REPAIR, INGUINAL, WITH LAPAROSCOPIC GROIN EXPLORATION, 5MM, GENSURG	29 (2.3%)	0 (0.0%)	HERNIA REPAIR, INGUINAL/HYDROCELECTOMY, BILATERAL, GENSURG	18 (1.4%)	1 (0.2%)
HERNIA REPAIR, INGUINAL/HYDROCELECTOMY, BILATERAL, GU	4 (0.3%)	0 (0.0%)	HERNIA REPAIR, INGUINAL/HYDROCELECTOMY, UNILATERAL, GENSURG	24 (1.9%)	4 (0.9%)
HERNIA REPAIR, INGUINAL/HYDROCELECTOMY, UNILATERAL, GU	18 (1.4%)	1 (0.2%)	HERNIA REPAIR, UMBILICAL, GENSURG	3 (0.2%)	0 (0.0%)
HERNIA REPAIR, VENTRAL, GENSURG	7 (0.5%)	1 (0.2%)	HIP, ANTERIOR ARTHROTOMY OSTEOPLASTY AND ARTHROSCOPY, ORTHO	2 (0.2%)	0 (0.0%)
HIP, ANTERIOR ARTHROTOMY OSTEOPLASTY, ORTHO	5 (0.4%)	0 (0.0%)	HIP, ARTHROGRAM, ORTHO	9 (0.7%)	1 (0.2%)
HIP, ARTHROSCOPY PROCEDURES, LABRAL REPAIR/OSTEOPLASTY, ORTHO	82 (6.4%)	3 (0.7%)	HIP, ARTHROSCOPY PROCEDURES, REVISION, LABRAL RECONSTRUCTION, ORTHO	30 (2.4%)	2 (0.5%)
HIP, ARTHROSCOPY, DIAGNOSTIC, ORTHO	6 (0.5%)	1 (0.2%)	HIP, CAPSULECTOMY/CAPSULOTOMY, ORTHO	1 (0.1%)	0 (0.0%)
HIP, CLOSED REDUCTION, DEVELOPMENTAL DISLOCATION, ORTHO	31 (2.4%)	1 (0.2%)	HIP, HARDWARE REMOVAL ILIAC SCREW, ORTHO	8 (0.6%)	1 (0.2%)
HIP, MODIFIED DUNN SUBCAPITAL OSTEOTOMY, ORTHO	3 (0.2%)	0 (0.0%)	HIP, OPEN EXPLORATION, ORTHO	2 (0.2%)	0 (0.0%)
HIP, OPEN REDUCTION, DEVELOPMENTAL DISLOCATION WITH OSTEOTOMY, ORTHO	7 (0.5%)	0 (0.0%)	HIP, OPEN REDUCTION, DEVELOPMENTAL DISLOCATION WITH PELVIC AND FEMORAL OSTEOTOMY, ORTHO	55 (4.3%)	9 (2.0%)
HIP, OPEN REDUCTION, DEVELOPMENTAL DISLOCATION, ORTHO	33 (2.6%)	3 (0.7%)	HIP, PAO AND ARTHROSCOPY, ORTHO	92 (7.2%)	12 (2.7%)
HIP, PAO AND FEMORAL OSTEOTOMY, ORTHO	23 (1.8%)	3 (0.7%)	HIP, PAO AND HARDWARE REMOVAL OPPOSITE HIP, ORTHO	11 (0.9%)	4 (0.9%)
HIP, PAO AND SURGICAL HIP DISLOCATION, ORTHO	12 (0.9%)	1 (0.2%)	HIP, PAO, ORTHO	298 (23.4%)	44 (10.0%)
HIP, PERCUTANEOUS TENDON RELEASE, ORTHO	19 (1.5%)	1 (0.2%)	HIP, SCFE IN SITU PINNING, ORTHO	4 (0.3%)	1 (0.2%)
HIP, SURGICAL HIP DISLOCATION, ORTHO	66 (5.2%)	6 (1.4%)	HIP, TENDON LENGTHENING/RELEASE, ORTHO	18 (1.4%)	0 (0.0%)
HUMERUS DISTAL, CLOSED REDUCTION, PERCUTANEOUS PINNING, ORTHO	1 (0.1%)	0 (0.0%)	HUMERUS DISTAL, ORIF, ORTHO	5 (0.4%)	0 (0.0%)
HUMERUS PROXIMAL, ORIF, ORTHO	13 (1.0%)	0 (0.0%)	HUMERUS, LENGTHENING, EXFIX, ORTHO	2 (0.2%)	1 (0.2%)
HUMERUS, LENGTHENING, MAGNETIC ROD, ORTHO	2 (0.2%)	0 (0.0%)	HUMERUS, OSTEOTOMY, DISTAL, ORTHO	14 (1.1%)	2 (0.5%)
HYDROCELECTOMY, GENSURG	2 (0.2%)	0 (0.0%)	HYDROCELECTOMY, GU	1 (0.1%)	0 (0.0%)
HYFRECATION OF LASHES, EYE	1 (0.1%)	0 (0.0%)	HYMENECTOMY, GYN	1 (0.1%)	0 (0.0%)
HYPOSPADIAS REPAIR FIRST STAGE, GU	47 (3.7%)	6 (1.4%)	HYPOSPADIAS REPAIR SECOND STAGE, GU	60 (4.7%)	11 (2.5%)
HYPOSPADIAS REPAIR WITH BUCCAL MUCOSA GRAFT, GU	40 (3.1%)	7 (1.6%)	HYPOSPADIAS REPAIR, DISTAL, GU	19 (1.5%)	0 (0.0%)
HYPOSPADIAS REPAIR, MEATAL ADVANCEMENT, GU	3 (0.2%)	0 (0.0%)	HYPOSPADIAS REPAIR, MIDSHAFT, GU	15 (1.2%)	0 (0.0%)
HYPOSPADIAS REPAIR, PROXIMAL, GU	24 (1.9%)	2 (0.5%)	ICP MONITOR INSERTION, NEURO	7 (0.5%)	2 (0.5%)
ICP MONITOR INSERTION/RM 25, NEURO	1 (0.1%)	0 (0.0%)	ILEOCECAL RESECTION, LAPAROSCOPIC ASSISTED, 5MM, GENSURG	39 (3.1%)	7 (1.6%)
ILEOCECAL RESECTION, LAPAROSCOPIC ASSISTED, ¿ 25 KG, GENSURG	3 (0.2%)	0 (0.0%)	ILEOSCOPY THROUGH STOMA, GI	2 (0.2%)	0 (0.0%)
ILEOSTOMY, GENSURG	5 (0.4%)	0 (0.0%)	ILIAC CREST BONE GRAFT, ORTHO	3 (0.2%)	0 (0.0%)
IMPEDANCE PROBE, GI	2 (0.2%)	0 (0.0%)	INCISION AND DRAINAGE, ABSCESS, GENSURG	3 (0.2%)	0 (0.0%)
INCISION AND DRAINAGE, ORL	4 (0.3%)	1 (0.2%)	INCISION AND DRAINAGE, WOUND, PLASTICS	1 (0.1%)	1 (0.2%)
INJECTION, TRIGGER POINT, PAIN	1 (0.1%)	0 (0.0%)	INTERMAXILLARY FIXATION APPLICATION/REMOVAL, PLASTICS	1 (0.1%)	0 (0.0%)
INTERNAL FIXATION UPPER EXTREMITY LENGTHENING W MAGNETIC NAIL, ORTHO	2 (0.2%)	0 (0.0%)	INTERVENTIONAL RADIOLOGY PROCEDURE, RADIOLOGY	7 (0.5%)	0 (0.0%)
INTESTINAL DUPLICATION RESECTION, LAPAROSCOPIC, 5MM, GENSURG	2 (0.2%)	0 (0.0%)	INTUBATION, RADIOLOGY	1 (0.1%)	0 (0.0%)
IRRIGATION AND DEBRIDEMENT WITH VAC DRESSING CHANGE, GENSURG	1 (0.1%)	0 (0.0%)	IRRIGATION AND DEBRIDEMENT WOUND, ORTHO	4 (0.3%)	0 (0.0%)
IRRIGATION AND DEBRIDEMENT WOUND, SOFT-TISSUE, PLASTICS	3 (0.2%)	0 (0.0%)	IUD INSERTION, GYN	1 (0.1%)	0 (0.0%)
KASAI PROCEDURE, GENSURG	4 (0.3%)	0 (0.0%)	KNEE EXPLORATION, OPEN, ORTHO	9 (0.7%)	0 (0.0%)
KNEE MANIPULATION, ORTHO	6 (0.5%)	0 (0.0%)	KNEE, ARTHROSCOPY ACL RECONSTRUCTION WITH BTB, ORTHO	21 (1.6%)	8 (1.8%)
KNEE, ARTHROSCOPY ACL RECONSTRUCTION WITH HAMSTRING TENDON, ORTHO	37 (2.9%)	4 (0.9%)	KNEE, ARTHROSCOPY ACL RECONSTRUCTION WITH IT BAND, ORTHO	99 (7.8%)	3 (0.7%)
KNEE, ARTHROSCOPY ACL RECONSTRUCTION WITH IT BAND/SATURN, ORTHO	2 (0.2%)	1 (0.2%)	KNEE, ARTHROSCOPY ACL RECONSTRUCTION WITH QUAD TENDON, ORTHO	2 (0.2%)	1 (0.2%)
KNEE, ARTHROSCOPY ACL REVISION WITH BTB, ORTHO	2 (0.2%)	0 (0.0%)	KNEE, ARTHROSCOPY FOR TIBIAL SPINE FRACTURE, ORTHO	10 (0.8%)	1 (0.2%)
KNEE, ARTHROSCOPY MCL RECONSTRUCTION, ORTHO	1 (0.1%)	0 (0.0%)	KNEE, ARTHROSCOPY PCL RECONSTRUCTION, ORTHO	5 (0.4%)	0 (0.0%)
KNEE, ARTHROSCOPY TO BE DONE IN IR, ORTHO	1 (0.1%)	0 (0.0%)	KNEE, ARTHROSCOPY WITH CARTILAGE BIOPSY, ORTHO	2 (0.2%)	0 (0.0%)
KNEE, ARTHROSCOPY WITH MENISCAL TRANSPLANT, ORTHO	2 (0.2%)	1 (0.2%)	KNEE, ARTHROSCOPY WITH MPFL RECONSTRUCTION, ORTHO	25 (2.0%)	9 (2.0%)
KNEE, ARTHROSCOPY WITH MPFL REPAIR, ORTHO	51 (4.0%)	0 (0.0%)	KNEE, ARTHROSCOPY WITH OATS PROCEDURE, ORTHO	2 (0.2%)	1 (0.2%)
KNEE, ARTHROSCOPY WITH OPEN EXPLORATION/REPAIR, ORTHO	12 (0.9%)	0 (0.0%)	KNEE, ARTHROSCOPY WITH OPEN MEDIAL PLICATION, ORTHO	14 (1.1%)	1 (0.2%)

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)
KNEE, ARTHROSCOPY WITH ROUX-GOLTHWAITHE, ORTHO	13 (1.0%)	1 (0.2%)	KNEE, ARTHROSCOPY WITH TIBIAL TUBERCLE OSTEOTOMY, ORTHO	94 (7.4%)	14 (3.2%)
KNEE, ARTHROSCOPY, MENISCUS DEBRIDEMENT/REPAIR, ORTHO	11 (0.9%)	2 (0.5%)	KNEE, ARTHROSCOPY, OCD DRILLING/FIXATION, ORTHO	2 (0.2%)	1 (0.2%)
KNEE, ARTHROSCOPY, ORTHO	62 (4.9%)	4 (0.9%)	KNEE, HARDWARE REMOVAL, ORTHO	11 (0.9%)	0 (0.0%)
KNEE, OPEN REDUCTION, CONGENITAL DISLOCATION, ORTHO	2 (0.2%)	0 (0.0%)	KNEE, OPEN REDUCTION, CONGENITAL PATELLAR DISLOCATION, ORTHO	2 (0.2%)	1 (0.2%)
KNEE, ORIF, PATELLA FRACTURE, ORTHO	2 (0.2%)	1 (0.2%)	KNEE, PATELLAR TENDON REPAIR, ORTHO	8 (0.6%)	0 (0.0%)
KNEE, TENDON LENGTHENING/RELEASE, ORTHO	12 (0.9%)	1 (0.2%)	KNEE, TENDON TRANSFER, ORTHO	1 (0.1%)	0 (0.0%)
KNEE, TIBIAL TUBERCLE FRACTURE, ORIF, ORTHO	3 (0.2%)	0 (0.0%)	KNEE, VERICEL IMPLANT, ORTHO	4 (0.3%)	1 (0.2%)
LABIAPLASTY, GYN	1 (0.1%)	0 (0.0%)	LACRIMAL DUCT PROBE, NO TUBES, EYE	5 (0.4%)	0 (0.0%)
LADD'S PROCEDURE, LAPAROSCOPIC, 5MM, GENSURG	5 (0.4%)	1 (0.2%)	LADD'S PROCEDURE, OPEN, GENSURG	4 (0.3%)	0 (0.0%)
LAMINECTOMY, DECOMPRESSION/EXPLORATION, NEURO	30 (2.4%)	3 (0.7%)	LAPAROSCOPIC HEMI-HYSTERECTOMY WITH MORCELLATOR, GYN	1 (0.1%)	0 (0.0%)
LAPAROSCOPIC HYSTERECTOMY, GYN	3 (0.2%)	0 (0.0%)	LAPAROSCOPIC NEPHRECTOMY, GU	8 (0.6%)	2 (0.5%)
LAPAROSCOPIC NEPHROURETERECTOMY, GU	1 (0.1%)	0 (0.0%)	LAPAROSCOPIC OOPHORECTOMY FOR FERTILITY PRESERVATION, GYN	11 (0.9%)	4 (0.9%)
LAPAROSCOPIC ORCHIDOPEXY 1ST STAGE, GU	1 (0.1%)	0 (0.0%)	LAPAROSCOPIC ORCHIDOPEXY 2ND STAGE, GU	9 (0.7%)	0 (0.0%)
LAPAROSCOPIC ORCHIDOPEXY, GU	4 (0.3%)	1 (0.2%)	LAPAROSCOPY, DIAGNOSTIC FOR OVARIAN TORSION, 3MM, GENSURG	1 (0.1%)	0 (0.0%)
LAPAROSCOPY, DIAGNOSTIC, 3MM, GENSURG	2 (0.2%)	0 (0.0%)	LAPAROSCOPY, DIAGNOSTIC, 5MM, GENSURG	36 (2.8%)	1 (0.2%)
LAPAROSCOPY, DIAGNOSTIC, GU	2 (0.2%)	0 (0.0%)	LAPAROSCOPY, DIAGNOSTIC, WITH ORCHIDOPEXY, GU	2 (0.2%)	0 (0.0%)
LAPAROSCOPY, SHORT, GYN	7 (0.5%)	2 (0.5%)	LAPAROTOMY, EXPLORATORY, 9 TO 25 KG, GENSURG	92 (7.2%)	1 (0.2%)
LAPAROTOMY, EXPLORATORY, ¡ 9 KG, GENSURG	4 (0.3%)	0 (0.0%)	LAPAROTOMY, EXPLORATORY, ¿ 25 KG, GENSURG	18 (1.4%)	3 (0.7%)
LAPAROTOMY, EXPLORATORY, GU	1 (0.1%)	1 (0.2%)	LAPAROTOMY, EXPLORATORY, GYN	32 (2.5%)	4 (0.9%)
LAPAROTOMY, STAGING OF OVARIAN MASS, GYN	10 (0.8%)	9 (2.0%)	LAPAROTOMY/ENDOSCOPY FOR BLUE RUBBER BLEB NEVUS SYNDROME, 9 TO 25 KG, GENSURG	1 (0.1%)	0 (0.0%)
LARYNGEAL CLEFT, ENDOSCOPIC REPAIR WITH LASER, ORL	267 (21.0%)	20 (4.5%)	LARYNGEAL CLEFT, OPEN REPAIR, ORL	1 (0.1%)	0 (0.0%)
LARYNGOPLASTY WITH IMPLANT, ORL	3 (0.2%)	0 (0.0%)	LARYNGOTRACHEAL RECONSTRUCTION, ORL	3 (0.2%)	0 (0.0%)
LASER ABLATION, CLEARPOINT, NEURO	2 (0.2%)	1 (0.2%)	LASER ABLATION, ROBOT ASSISTED, NEURO	10 (0.8%)	3 (0.7%)
LEG, AMPUTATION, ORTHO	12 (0.9%)	1 (0.2%)	LEG, OPEN EXERTIONAL COMPARTMENT RELEASE, ORTHO	2 (0.2%)	0 (0.0%)
LEG, TENDON LENGTHENING, ORTHO	1 (0.1%)	0 (0.0%)	LEG/FOOT, HARDWARE REMOVAL, ORTHO	10 (0.8%)	2 (0.5%)
LENSECTOMY, EYE	1 (0.1%)	0 (0.0%)	LIPOSUCTION, PLASTICS	2 (0.2%)	1 (0.2%)
LIPOSUCTION/SUCTION ASSISTED LIPECTOMY, PLASTICS	8 (0.6%)	1 (0.2%)	LIVER TRANSPLANT RECIPIENT, LIVING DONOR, GENSURG	3 (0.2%)	0 (0.0%)
LIVER TRANSPLANT, GENSURG	1 (0.1%)	0 (0.0%)	LOCAL/REGIONAL FLAP, PLASTICS	5 (0.4%)	0 (0.0%)
LOWER EXTREMITY, EPIPHYSIODESIS, ORTHO	93 (7.3%)	3 (0.7%)	LOWER EXTREMITY, HARDWARE REMOVAL EXPOSED PIN, ORTHO	1 (0.1%)	0 (0.0%)
LOWER EXTREMITY, HEMIEPIPHYSIODESIS, ORTHO	13 (1.0%)	5 (1.1%)	LOWER EXTREMITY, INTERNAL LENGTHENING WITH MAGNETIC NAIL, ORTHO	2 (0.2%)	1 (0.2%)
LOWER EXTREMITY, PERIPHERAL NERVE NEUROPLASTY/REPAIR/NERVE GRAFTING, ORTHO	2 (0.2%)	0 (0.0%)	LP SHUNT INSERTION/REVISION/REMOVAL, NEURO	2 (0.2%)	0 (0.0%)
LUMBAR DRAIN PLACEMENT, NEURO	1 (0.1%)	0 (0.0%)	LUMBAR PUNCTURE FOR INTRATHECAL MEDICATION, NEUROLOGY	1 (0.1%)	0 (0.0%)
LUMBAR PUNCTURE WITH OR WITHOUT INTRATHECAL MEDICATION, ONCOLOGY	21 (1.6%)	0 (0.0%)	LUMBAR PUNCTURE, NEURO	2 (0.2%)	0 (0.0%)
LUMBAR PUNCTURE, NEUROLOGY	9 (0.7%)	0 (0.0%)	LUMBAR PUNCTURE, ONCOLOGY	2 (0.2%)	0 (0.0%)
LUMBAR PUNCTURE, PAIN	4 (0.3%)	3 (0.7%)	LUNG LOBECTOMY, THORACOSCOPIC, 3MM, GENSURG	8 (0.6%)	2 (0.5%)
LUNG LOBECTOMY, THORACOSCOPIC, 5MM, GENSURG	20 (1.6%)	0 (0.0%)	LUNG WEDGE RESECTION, THORACOSCOPIC, 5MM, GENSURG	15 (1.2%)	3 (0.7%)
LYSIS OF ADHESIONS, GENSURG	1 (0.1%)	0 (0.0%)	MACROSTOMIA REPAIR, REVISION ORAL COMMISURE, PLASTICS	1 (0.1%)	0 (0.0%)
MALAR IMPLANT, PLASTICS	1 (0.1%)	0 (0.0%)	MALFORMATION EXCISION (LM/VM), LARGE, GENSURG	3 (0.2%)	0 (0.0%)
MALFORMATION EXCISION (LM/VM), MEDIUM, GENSURG	14 (1.1%)	1 (0.2%)	MALFORMATION EXCISION (LM/VM), SMALL, GENSURG	6 (0.5%)	0 (0.0%)
MANDIBLE FRACTURE, CLOSED REDUCTION ¡1 HOUR, PLASTICS	1 (0.1%)	0 (0.0%)	MANDIBLE FRACTURE, ORIF ¿1.5 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)
MANDIBULAR RESECTION ¡2 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)	MANDIBULAR RESECTION ¿2 HOURS, PLASTICS	5 (0.4%)	0 (0.0%)
MANDIBULAR/MAXILLARY RECONSTRUCTION ¡3.5 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)	MANIPULATION JOINT, ORTHO	1 (0.1%)	0 (0.0%)
MASS EXCISION, BASIC, ORL	15 (1.2%)	0 (0.0%)	MASS EXCISION, COMPLEX, ORL	21 (1.6%)	5 (1.1%)
MASTECTOMY, PLASTICS	3 (0.2%)	0 (0.0%)	MASTOPEXY BILATERAL, PLASTICS	1 (0.1%)	0 (0.0%)
MASTOPEXY UNILATERAL, PLASTICS	3 (0.2%)	0 (0.0%)	MEATOPLASTY, GU	1 (0.1%)	0 (0.0%)
MEATOTOMY, URETHRAL, GU	1 (0.1%)	0 (0.0%)	MEDIASTINAL MASS RESECTION, THORACOSCOPIC, 5MM, GENSURG	16 (1.3%)	3 (0.7%)
MEGAURETER REPAIR WITH CYSTOSCOPY, GU	4 (0.3%)	0 (0.0%)	MEGAURETER REPAIR, GU	3 (0.2%)	2 (0.5%)
METOIDIOPLASTY, GU	2 (0.2%)	0 (0.0%)	MICRODIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY WITH LASER, ORL	4 (0.3%)	0 (0.0%)
MICRODIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY WITH SUSPENSION, ORL	9 (0.7%)	4 (0.9%)	MICRODIRECT LARYNGOSCOPY AND/OR BRONCHOSCOPY, ORL	1 (0.1%)	0 (0.0%)
MICRODISCECTOMY, NEURO	69 (5.4%)	4 (0.9%)	MICRODISCECTOMY/RM25, NEURO	3 (0.2%)	0 (0.0%)
MICROTIA CONSTRUCTION, AUTOLOGOUS, 1ST STAGE, PLASTICS	7 (0.5%)	0 (0.0%)	MICROTIA CONSTRUCTION, WITH IMPLANT, PLASTICS	1 (0.1%)	2 (0.5%)
MICROTIA RECONSTRUCTION 3RD STAGE, PLASTICS	1 (0.1%)	0 (0.0%)	MONTI, CATHETERIZABLE CHANNEL, GU	13 (1.0%)	2 (0.5%)
MRT SCANNING, NEURO	2 (0.2%)	2 (0.5%)	MUSCLE BIOPSY, ORTHO	2 (0.2%)	0 (0.0%)
MYELOMENINGOCELE REPAIR, NEURO	4 (0.3%)	0 (0.0%)	MYRINGOPLASTY, GRAFT, ORL	1 (0.1%)	0 (0.0%)
MYRINGOPLASTY, PATCH, ORL	4 (0.3%)	0 (0.0%)	NASAL CAUTERIZATION, ORL	4 (0.3%)	0 (0.0%)
NASAL ENDOSCOPY, ORL	10 (0.8%)	0 (0.0%)	NASAL RECONSTRUCTION, PLASTICS	4 (0.3%)	0 (0.0%)
NASAL/CILIARY BIOPSY, ORL	3 (0.2%)	0 (0.0%)	NASOPHARYNGEAL ANGIOFIBROMA RESECTION, ENDOSCOPIC AND OPEN APPROACH, ORL	1 (0.1%)	0 (0.0%)
NASOPHARYNGEAL ANGIOFIBROMA RESECTION, ENDOSCOPIC, ORL	2 (0.2%)	0 (0.0%)	NECK MASS EXCISION, ORL	22 (1.7%)	0 (0.0%)
NEPHRECTOMY, LAPAROSCOPIC, 5MM, GENSURG	16 (1.3%)	0 (0.0%)	NEPHRECTOMY, PARTIAL, GU	5 (0.4%)	0 (0.0%)
NEPHRECTOMY, PARTIAL/TOTAL, OPEN, GENSURG	23 (1.8%)	8 (1.8%)	NEPHRECTOMY, SIMPLE, GU	1 (0.1%)	0 (0.0%)
NEPHRECTOMY, WILMS TUMOR, GENSURG	13 (1.0%)	0 (0.0%)	NEPHROLITHOTOMY, GU	1 (0.1%)	0 (0.0%)
NEPHROURETERECTOMY, GENSURG	9 (0.7%)	0 (0.0%)	NEPHROURETERECTOMY, GU	1 (0.1%)	0 (0.0%)
NERVE REPAIR, PLASTICS	2 (0.2%)	0 (0.0%)	NISSEN FUNDOPLICATION, LAPAROSCOPIC, 5MM, GENSURG	11 (0.9%)	0 (0.0%)
NISSEN FUNDOPLICATION, LAPAROSCOPIC, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	NISSEN FUNDOPLICATION, OPEN, GENSURG	6 (0.5%)	0 (0.0%)
NISSEN FUNDOPLICATION, ROBOTIC, GENSURG	8 (0.6%)	0 (0.0%)	NM PLUS: ARTHROGRAM, ORTHO	3 (0.2%)	5 (1.1%)
NM PLUS: LOWER EXTREMITY, SOFT TISSUE LENGTHENING X2, ORTHO	2 (0.2%)	0 (0.0%)	NM PLUS: LOWER EXTREMITY, SOFT TISSUE LENGTHENING X4, ORTHO	2 (0.2%)	0 (0.0%)
NM PLUS: LOWER EXTREMITY, SOFT TISSUE LENGTHENING X6, ORTHO	1 (0.1%)	0 (0.0%)	NM PLUS: TIBIA, OSTEOTOMY, UNILATERAL, ORTHO	4 (0.3%)	0 (0.0%)
NM: FEMUR OSTEOTOMY, DISTAL, W/ PATELLAR TENDON ADVANCEMENT, ORTHO	4 (0.3%)	1 (0.2%)	NM: PELVIS, DEGA OSTEOTOMY, BILATERAL, W/ FEMUR OSTEOTOMY, PROXIMAL, BILATERAL, ORTHO	5 (0.4%)	4 (0.9%)
NM: PELVIS, DEGA OSTEOTOMY, UNILATERAL, W/ FEMUR OSTEOTOMY, PROXIMAL, BILATERAL, ORTHO	6 (0.5%)	1 (0.2%)	NM: PELVIS, DEGA OSTEOTOMY, UNILATERAL, W/ FEMUR OSTEOTOMY, PROXIMAL, UNILATERAL, ORTHO	1 (0.1%)	2 (0.5%)
ODONTECTOMY, PLASTICS	11 (0.9%)	0 (0.0%)	OMPHALOCELE FINAL CLOSURE, GENSURG	2 (0.2%)	0 (0.0%)
OMPHALOCELE INITIAL REPAIR, GENSURG	2 (0.2%)	1 (0.2%)	OMPHALOCELE STAGED CLOSURE, GENSURG	3 (0.2%)	0 (0.0%)
OOPHORECTOMY, LAPAROSCOPIC, 5MM, GENSURG	5 (0.4%)	0 (0.0%)	OPEN REDUCTION HIP WITH SPICA CAST, ORTHO	7 (0.5%)	0 (0.0%)
OPEN REDUCTION HIP, FEMORAL AND PELVIC OSTEOTOMY, ORTHO	3 (0.2%)	0 (0.0%)	OPEN REDUCTION HIP, FEMORAL OSTEOTOMY WITH SPICA CAST, ORTHO	5 (0.4%)	0 (0.0%)
OPEN SEPTO-RHINOPLASTY, PLASTICS	3 (0.2%)	1 (0.2%)	ORAL CAVITY LESION COBLATION, ORL	1 (0.1%)	1 (0.2%)
ORAL CAVITY LESION EXCISION/BIOPSY, ORL	3 (0.2%)	1 (0.2%)	ORAL LESION EXCISION, ORL	3 (0.2%)	0 (0.0%)
ORBIT TUMOR, EYE	3 (0.2%)	0 (0.0%)	ORCHIDOPEXY, BILATERAL, GU	11 (0.9%)	1 (0.2%)
ORCHIDOPEXY, OPEN, GENSURG	4 (0.3%)	0 (0.0%)	ORCHIDOPEXY, UNILATERAL, GU	18 (1.4%)	0 (0.0%)
ORCHIDOPEXY, WITH INGUINAL HERNIA REPAIR, GU	1 (0.1%)	0 (0.0%)	ORCHIECTOMY, RADICAL, GU	3 (0.2%)	0 (0.0%)
ORCHIECTOMY, SIMPLE, GU	1 (0.1%)	0 (0.0%)	ORIF BLOWOUT ORBIT FRACTURE, PLASTICS	8 (0.6%)	0 (0.0%)
ORIF CALCANEUS, ORTHO	2 (0.2%)	0 (0.0%)	ORIF FEMUR DISTAL, ORTHO	2 (0.2%)	0 (0.0%)
ORIF FEMUR PROXIMAL, ORTHO	4 (0.3%)	0 (0.0%)	ORIF FRONTAL SINUS FRACTURE, PLASTICS	1 (0.1%)	0 (0.0%)
ORIF HAND, PLASTICS	1 (0.1%)	0 (0.0%)	ORIF HUMERAL CONDYLE, ORTHO	2 (0.2%)	0 (0.0%)
ORIF INTERTROCHANTERIC, ORTHO	9 (0.7%)	0 (0.0%)	ORIF MONTEGGIA FRACTURE, ORTHO	1 (0.1%)	0 (0.0%)
ORIF RADIUS AND/OR ULNA WITH IM ROD, ORTHO	2 (0.2%)	0 (0.0%)	ORIF TIBIA WITH IM ROD, ORTHO	3 (0.2%)	0 (0.0%)
ORIF ZYGOMATIC FRACTURE, PLASTICS	6 (0.5%)	1 (0.2%)	ORO-NASAL FISTULA REPAIR, BILATERAL (WITH JBM), PLASTICS	63 (4.9%)	0 (0.0%)
ORO-NASAL FISTULA REPAIR, BILATERAL (WITHOUT JBM), PLASTICS	92 (7.2%)	0 (0.0%)	ORO-NASAL FISTULA REPAIR, UNILATERAL (WITH JBM), PLASTICS	13 (1.0%)	4 (0.9%)
ORO-NASAL FISTULA REPAIR, UNILATERAL (WITHOUT JBM), PLASTICS	13 (1.0%)	4 (0.9%)	ORO-NASAL FISTULA REPAIR, WITH PRE MAXIALLARY OSTEOTOMY (WITHOUT JBM), PLASTICS	16 (1.3%)	0 (0.0%)
ORTHOGNATHIC, BIMAXILLARY OSTEOTOMY ¡4 HOURS, PLASTICS	36 (2.8%)	1 (0.2%)	ORTHOGNATHIC, BIMAXILLARY OSTEOTOMY ¿4 HOURS, PLASTICS	6 (0.5%)	0 (0.0%)
ORTHOGNATHIC, CRANIOFACIAL/COMPLEX ¿4 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)	ORTHOGNATHIC, GENIOPLASTY, PLASTICS	1 (0.1%)	0 (0.0%)
ORTHOGNATHIC, LE FORT I ¡2 HOURS, PLASTICS	17 (1.3%)	5 (1.1%)	ORTHOGNATHIC, LE FORT I ¿2 HOURS, PLASTICS	174 (13.7%)	5 (1.1%)
ORTHOGNATHIC, LE FORT III, PLASTICS	5 (0.4%)	0 (0.0%)	ORTHOGNATHIC, SAGITTAL SPLIT OSTEOTOMY, PLASTICS	36 (2.8%)	5 (1.1%)
ORTHOGNATHIC, SURGICALLY ASSISTED MAXILLARY EXPANSION, PLASTICS	6 (0.5%)	1 (0.2%)	OSSEOINTEGRATED AUDITORY IMPLANT, TRANSCUTANEOUS, ORL	1 (0.1%)	0 (0.0%)
OSSICULAR CHAIN RECONSTRUCTION, ORL	2 (0.2%)	0 (0.0%)	OSTEOCHONDROMA EXCISION, COMPLEX, ORTHO	29 (2.3%)	2 (0.5%)
OSTEOCHONDROMA EXCISION, SIMPLE, ORTHO	17 (1.3%)	2 (0.5%)	OSTEOID OSTEOMA RESECTION, NEURO	10 (0.8%)	1 (0.2%)
OSTEOTOMY ANKLE AND/OR FOOT, ORTHO	19 (1.5%)	0 (0.0%)	OSTEOTOMY FEMUR ADULT, BILATERAL, ORTHO	7 (0.5%)	0 (0.0%)
OSTEOTOMY FEMUR CHILD, BILATERAL WITH CAST, ORTHO	1 (0.1%)	0 (0.0%)	OSTEOTOMY FEMUR CHILD, UNILATERAL WITH CAST, ORTHO	3 (0.2%)	0 (0.0%)
OSTEOTOMY FEMUR CHILD, UNILATERAL, ORTHO	18 (1.4%)	0 (0.0%)	OSTEOTOMY FEMUR VARUS/VALGUS DEROTATIONAL BILATERAL, ORTHO	5 (0.4%)	0 (0.0%)
OSTEOTOMY FEMUR VARUS/VALGUS DEROTATIONAL, UNILATERAL, ORTHO	7 (0.5%)	0 (0.0%)	OSTEOTOMY FINGER/HAND, PLASTICS	12 (0.9%)	2 (0.5%)
OSTEOTOMY PELVIS CHIARI BILATERAL WITH FEMORAL OSTEOTOMY BILATERAL, ORTHO	1 (0.1%)	0 (0.0%)	OSTEOTOMY PELVIS CHIARI UNILATERAL WITH FEMORAL OSTEOTOMY BILATERAL, ORTHO	4 (0.3%)	0 (0.0%)
OSTEOTOMY PELVIS CHIARI UNILATERAL WITH FEMORAL OSTEOTOMY UNILATERAL, ORTHO	2 (0.2%)	0 (0.0%)	OSTEOTOMY PELVIS DEGA PEMBERTON INNOMINATE BILATERAL WITH FEMORAL OSTEOTOMY UNILATERAL, ORTHO	3 (0.2%)	0 (0.0%)
OSTEOTOMY PELVIS DEGA PEMBERTON INNOMINATE UNILATERAL WITH FEMORAL OSTEOTOMY BILATERAL, ORTHO	21 (1.6%)	0 (0.0%)	OSTEOTOMY PELVIS SHELF BILATERAL WITH FEMORAL OSTEOTOMY UNILATERAL, ORTHO	2 (0.2%)	0 (0.0%)
OSTOMY CLOSURE, 9 TO 25 KG, GENSURG	66 (5.2%)	3 (0.7%)	OSTOMY CLOSURE, ¡ 9 KG, GENSURG	11 (0.9%)	2 (0.5%)
OSTOMY CLOSURE, ¿ 25 KG, GENSURG	19 (1.5%)	1 (0.2%)	OSTOMY REVISION, 9 TO 25 KG, GENSURG	9 (0.7%)	3 (0.7%)
OSTOMY REVISION, ¡ 9 KG, GENSURG	2 (0.2%)	0 (0.0%)	OSTOMY REVISION, ¿ 25 KG, GENSURG	1 (0.1%)	5 (1.1%)
OSTOMY, 9 TO 25 KG, GENSURG	6 (0.5%)	0 (0.0%)	PALATAL FISTULA CLOSURE, PLASTICS	4 (0.3%)	0 (0.0%)
PALATOPLASTY, FURLOW, PLASTICS	36 (2.8%)	3 (0.7%)	PANCREATECTOMY, DISTAL, OPEN, GENSURG	3 (0.2%)	0 (0.0%)

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)
PANCREATECTOMY, WHIPPLE, OPEN, GENSURG	1 (0.1%)	0 (0.0%)	PANNICULECTOMY/ABDOMINOPLASTY, PLASTICS	7 (0.5%)	1 (0.2%)
PAO HIP WITH SURGICAL DISLOCATION AND FEMORAL OSTEOTOMY, ORTHO	1 (0.1%)	0 (0.0%)	PARATHYROIDECTOMY, GENSURG	3 (0.2%)	0 (0.0%)
PAROTIDECTOMY, ORL	19 (1.5%)	1 (0.2%)	PATELLA REALIGNMENT, ORTHO	8 (0.6%)	0 (0.0%)
PATELLA TENDON SHORTENING, ORTHO	1 (0.1%)	0 (0.0%)	PECTUS CARINATUM REPAIR, OPEN, GENSURG	2 (0.2%)	0 (0.0%)
PECTUS EXCAVATUM REPAIR, NUSS, THORACOSCOPIC, GENSURG	32 (2.5%)	1 (0.2%)	PECTUS EXCAVATUM REPAIR, OPEN, GENSURG	9 (0.7%)	0 (0.0%)
PECTUS STRUT REMOVAL, ATKINS, GENSURG	1 (0.1%)	1 (0.2%)	PECTUS STRUT REMOVAL, NUSS, GENSURG	3 (0.2%)	0 (0.0%)
PEG INSERTION, GENSURG	12 (0.9%)	1 (0.2%)	PELVIS, BERNESE TRIPLE OSTEOTOMY, ORTHO	3 (0.2%)	1 (0.2%)
PELVIS, BILATERAL ILIAC OSTEOTOMY FOR EXSTROPHY, ANTERIOR, ORTHO	1 (0.1%)	0 (0.0%)	PELVIS, CHIARI OSTEOTOMY, ORTHO	1 (0.1%)	0 (0.0%)
PELVIS, DEGA PEMBERTON INNOMINATE OSTEOTOMY AND FEMORAL OSTEOTOMY BILATERAL, ORTHO	59 (4.6%)	1 (0.2%)	PELVIS, DEGA PEMBERTON INNOMINATE OSTEOTOMY AND FEMORAL OSTEOTOMY UNILATERAL, ORTHO	5 (0.4%)	0 (0.0%)
PELVIS, DEGA PEMBERTON INNOMINATE OSTEOTOMY, ORTHO	10 (0.8%)	0 (0.0%)	PELVIS, ISCHIAL AVULSION FRACTURE REPAIR, ORTHO	9 (0.7%)	0 (0.0%)
PELVIS/ACETABULUM, ORIF, ORTHO	4 (0.3%)	0 (0.0%)	PENILE LYMPHEDEMA EXCISION WITH PENILE RECONSTRUCTION, GENSURG	1 (0.1%)	0 (0.0%)
PERCUTANEOUS NEPHROLITHOTOMY WITH HOLMIUM LASER, GU	22 (1.7%)	2 (0.5%)	PERCUTANEOUS NEPHROLITHOTOMY WITH SHOCK PULSE, GU	4 (0.3%)	0 (0.0%)
PERIPHERAL ARTERY RECONSTRUCTION, GENSURG	4 (0.3%)	0 (0.0%)	PERITONEAL DIALYSIS CATHETER INSERTION, GENSURG	6 (0.5%)	0 (0.0%)
PERITONEAL DIALYSIS CATHETER REMOVAL, GENSURG	2 (0.2%)	0 (0.0%)	PHALLOPLASTY, ANTEROLATERAL FLAP, NON MICROSURGICAL, PLASTICS	1 (0.1%)	0 (0.0%)
PHALLOPLASTY, RFFF, PLASTICS	17 (1.3%)	0 (0.0%)	PHARYNGEAL FLAP, PLASTICS	94 (7.4%)	6 (1.4%)
PHERESIS CATHETER/HEMODIALYSIS CATHETER INSERTION, NOT TUNNELED, GENSURG	16 (1.3%)	0 (0.0%)	PHERESIS CATHETER/HEMODIALYSIS CATHETER INSERTION, TUNNELED, GENSURG	26 (2.0%)	0 (0.0%)
PHYSEAL ARREST LOWER EXTREMITY, ORTHO	7 (0.5%)	0 (0.0%)	PHYSEAL BAR EXCISION, ORTHO	4 (0.3%)	0 (0.0%)
PICC PLACEMENT, MEDICAL	59 (4.6%)	0 (0.0%)	PICC PLACEMENT, NON-SEDATED, MEDICAL	42 (3.3%)	0 (0.0%)
PILONIDAL SINUS EXCISION, GENSURG	1 (0.1%)	0 (0.0%)	PLEURODESIS WITH BLEB RESECTION, THORACOSCOPIC, GENSURG	7 (0.5%)	3 (0.7%)
POLLICIZATION, PLASTICS	2 (0.2%)	1 (0.2%)	POLYDACTYLY EXCISION, COMPLEX, PLASTICS	3 (0.2%)	0 (0.0%)
POLYDACTYLY EXCISION, SIMPLE, PLASTICS	2 (0.2%)	0 (0.0%)	PORT-A-CATH INSERTION/REVISION, GENSURG	96 (7.5%)	1 (0.2%)
PORT-A-CATH REMOVAL, GENSURG	4 (0.3%)	0 (0.0%)	PORT-A-CATH REVISION, GENSURG	10 (0.8%)	0 (0.0%)
PORTOCAVAL/SPLENORENAL SHUNT, GENSURG	12 (0.9%)	1 (0.2%)	POSTERIOR TRACHEOPEXY, OPEN, GENSURG	95 (7.5%)	3 (0.7%)
PREAURICULAR MASS/SINUS EXCISION, ORL	6 (0.5%)	1 (0.2%)	PROCTOPEXY, LAPAROSCOPIC, 5MM, GENSURG	2 (0.2%)	0 (0.0%)
PROCTOPEXY, LAPAROSCOPIC, ¿ 9 KG, GENSURG	1 (0.1%)	0 (0.0%)	PULL THROUGH FOR HIRSCHPRUNG'S, OPEN, GENSURG	1 (0.1%)	0 (0.0%)
PULL THROUGH, ILEOANAL WITH EEA, LAPAROSCOPIC ASSISTED, 5MM, GENSURG	8 (0.6%)	1 (0.2%)	PULL THROUGH, ILEOANAL WITH EEA, LAPAROSCOPIC ASSISTED, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)
PULL THROUGH, ILEOANAL WITH RECTAL MUCOSECTOMY, LAPAROSCOPIC ASSISTED, 5MM, GENSURG	12 (0.9%)	2 (0.5%)	PULL THROUGH, OPEN, 9 TO 25 KG, GENSURG	2 (0.2%)	3 (0.7%)
PULL THROUGH, OPEN, ¡ 9 KG, GENSURG	1 (0.1%)	1 (0.2%)	PULL THROUGH, OPEN, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)
PULL THROUGH, SOAVE, GENSURG	1 (0.1%)	0 (0.0%)	PULL THROUGH, TRANSANAL, LAPAROSCOPIC ASSISTED 3MM, GENSURG	2 (0.2%)	2 (0.5%)
PULL THROUGH, TRANSANAL, LAPAROSCOPIC ASSISTED 5MM, GENSURG	1 (0.1%)	0 (0.0%)	PULSE DYE LASER TREATMENT, DERM	1 (0.1%)	0 (0.0%)
PYELOPLASTY, COMPLEX, GU	3 (0.2%)	0 (0.0%)	PYELOPLASTY, GU	30 (2.4%)	5 (1.1%)
PYELOPLASTY/CYSTOSCOPY/RETROGRADE PYELOGRAM, GU	5 (0.4%)	0 (0.0%)	PYLOROPLASTY, GENSURG	2 (0.2%)	0 (0.0%)
RADIUS/ULNA, CLOSED REDUCTION, PERCUTANEOUS PINNING, ORTHO	1 (0.1%)	0 (0.0%)	RADIUS/ULNA, ORIF, ORTHO	20 (1.6%)	1 (0.2%)
RECTAL SIGMOID RESECTION, GENSURG	2 (0.2%)	0 (0.0%)	RECTOPEXY, OPEN, GENSURG	17 (1.3%)	1 (0.2%)
RECTOPEXY, ROBOTIC, GENSURG	4 (0.3%)	3 (0.7%)	REGIONAL BLOCK, PAIN	1 (0.1%)	0 (0.0%)
REMOVAL PROSTHESIS/PLACEMENT CEMENT SPACER, ORTHO	1 (0.1%)	0 (0.0%)	RENAL AUTOTRANSPLANT, GENSURG	16 (1.3%)	2 (0.5%)
RENAL TRANSPLANT/LIVING DONOR, RECIPIENT, GENSURG	23 (1.8%)	3 (0.7%)	RENAL TRANSPLANT/LIVING DONOR, RECIPIENT, LAPAROSCOPIC NEPHRECTOMY, GENSURG	4 (0.3%)	2 (0.5%)
RETINA, INSERTION SCLERAL BUCKLE, EYE	4 (0.3%)	0 (0.0%)	RETINA, VITRECTOMY, EYE	5 (0.4%)	1 (0.2%)
RETROPERITONEAL NODE DISSECTION, GU	4 (0.3%)	0 (0.0%)	RETROPERITONEAL TUMOR RESECTION, GENSURG	16 (1.3%)	6 (1.4%)
RHINOPLASTY, PLASTICS	8 (0.6%)	0 (0.0%)	RIB RESECTION, GENSURG	8 (0.6%)	0 (0.0%)
ROBOTIC ABDOMINAL PROCEDURE, GENSURG	6 (0.5%)	2 (0.5%)	ROBOTIC ARTIFICIAL URINARY SPHINCTER, GU	2 (0.2%)	0 (0.0%)
ROBOTIC BLADDER NECK SUSPENSION, GU	1 (0.1%)	0 (0.0%)	ROBOTIC CALYCEAL DIVERTICULECTOMY, GU	3 (0.2%)	0 (0.0%)
ROBOTIC CONTINENT STOMA, GU	3 (0.2%)	0 (0.0%)	ROBOTIC EXCISION OF RETROVESICAL STRUCTURE, GU	2 (0.2%)	0 (0.0%)
ROBOTIC EXCISION, GU	9 (0.7%)	2 (0.5%)	ROBOTIC EXTRAVESICAL URETERAL REIMPLANT, GU	9 (0.7%)	2 (0.5%)
ROBOTIC HYSTERECTOMY, GU	1 (0.1%)	0 (0.0%)	ROBOTIC INTRAVESICLE URETERAL REIMPLANT, BILATERAL, GU	1 (0.1%)	0 (0.0%)
ROBOTIC INTRAVESICLE URETERAL REIMPLANT, UNILATERAL, GU	2 (0.2%)	0 (0.0%)	ROBOTIC NEPHRECTOMY, GU	33 (2.6%)	7 (1.6%)
ROBOTIC NEPHRECTOMY, PARTIAL, GU	11 (0.9%)	3 (0.7%)	ROBOTIC OOPHORECTOMY, GU	1 (0.1%)	0 (0.0%)
ROBOTIC ORCHIDOPEXY, GU	1 (0.1%)	0 (0.0%)	ROBOTIC PYELOPLASTY, COMPLEX, GU	7 (0.5%)	0 (0.0%)
ROBOTIC PYELOPLASTY, GU	135 (10.6%)	26 (5.9%)	ROBOTIC PYELOPLASTY, REDO, GU	9 (0.7%)	0 (0.0%)
ROBOTIC RETROPERITONEAL NODE DISSECTION, GU	3 (0.2%)	1 (0.2%)	ROBOTIC URETEROLITHOTOMY, GU	1 (0.1%)	0 (0.0%)
ROBOTIC URETEROURETEROSTOMY, GU	16 (1.3%)	0 (0.0%)	ROBOTIC URINARY DIVERSION, GU	3 (0.2%)	0 (0.0%)
ROBOTIC VAGINAL PULL THROUGH, GU	1 (0.1%)	0 (0.0%)	ROBOTIC VAGINECTOMY, GENDER REASSIGNMENT, GU	6 (0.5%)	0 (0.0%)
ROTATIONPLASTY, ORTHO	7 (0.5%)	1 (0.2%)	SACRAL NEUROMODULATOR INSERTION, STAGE 1, GENSURG	2 (0.2%)	0 (0.0%)
SACROCOCCYGEAL TERATOMA RESECTION, GENSURG	4 (0.3%)	0 (0.0%)	SALIVARY GLAND DUCT EXPLORATION/REROUTING, ORL	5 (0.4%)	1 (0.2%)
SALIVARY GLANDS, BOTOX INJECTION, ORL	7 (0.5%)	2 (0.5%)	SALPINGO OOPHORECTOMY, GENSURG	4 (0.3%)	0 (0.0%)
SCAPHOID, ORIF, ORTHO	13 (1.0%)	0 (0.0%)	SCAPULA, ORIF, ORTHO	1 (0.1%)	0 (0.0%)
SCAPULECTOMY, ORTHO	2 (0.2%)	0 (0.0%)	SCAR REVISION BREAST/CHEST, PLASTICS	1 (0.1%)	0 (0.0%)
SCAR REVISION LARGE, PLASTICS	2 (0.2%)	0 (0.0%)	SCAR REVISION MEDIUM, PLASTICS	1 (0.1%)	0 (0.0%)
SCAR REVISION, GENSURG	1 (0.1%)	0 (0.0%)	SCLEROTHERAPY, GENSURG	3 (0.2%)	0 (0.0%)
SCROTAL EXPLORATION, TESTIS BIOPSY, GU	1 (0.1%)	0 (0.0%)	SCROTOPLASTY, GU	3 (0.2%)	0 (0.0%)
SECONDARY SURGEON	5 (0.4%)	2 (0.5%)	SEPTOPEXY, GU	1 (0.1%)	0 (0.0%)
SEPTOPLASTY, ORL	12 (0.9%)	2 (0.5%)	SEPTOPLASTY, PLASTICS	58 (4.6%)	5 (1.1%)
SHOULDER EXPLORATION, ORTHO	5 (0.4%)	0 (0.0%)	SHOULDER, ARTHROSCOPY CAPSULORRHAPHY, ORTHO	3 (0.2%)	1 (0.2%)
SHOULDER, ARTHROSCOPY WITH BANKART REPAIR, ORTHO	6 (0.5%)	0 (0.0%)	SHOULDER, ARTHROSCOPY, ORTHO	14 (1.1%)	1 (0.2%)
SHOULDER, CLAVICLE, EXCISION, ORTHO	1 (0.1%)	0 (0.0%)	SHOULDER, LATARJET RECONSTRUCTION, ORTHO	14 (1.1%)	0 (0.0%)
SHOULDER, OPEN BANKART REPAIR, ORTHO	6 (0.5%)	2 (0.5%)	SHOULDER, OPEN CAPSULORRHAPHY, ORTHO	14 (1.1%)	1 (0.2%)
SHOULDER, OSTEOTOMY, HUMERUS, PROXIMAL, ORTHO	13 (1.0%)	0 (0.0%)	SHOULDER, SCAPULA, EXCISION, ORTHO	1 (0.1%)	1 (0.2%)
SHOULDER, SCAPULA, OSTEOTOMY, ORTHO	2 (0.2%)	0 (0.0%)	SHOULDER, SCAPULA, RECONSTRUCTION, ORTHO	1 (0.1%)	0 (0.0%)
SHOULDER, SCAPULA, WOODWARD, ORTHO	9 (0.7%)	0 (0.0%)	SHOULDER, SCAPULOTHORACIC FUSION, ORTHO	3 (0.2%)	0 (0.0%)
SHOULDER, STERNOCLAVICULAR REPAIR, ORTHO	10 (0.8%)	0 (0.0%)	SHOULDER, TENDON LENGTHENING, ORTHO	3 (0.2%)	0 (0.0%)
SHOULDER, TENDON REPAIR, ORTHO	1 (0.1%)	0 (0.0%)	SHOULDER, TENDON TRANSFERS, BRACHIAL PLEXUS, ORTHO	10 (0.8%)	1 (0.2%)
SHOULDER, TENDON TRANSFERS, NON BRACHIAL PLEXUS, ORTHO	17 (1.3%)	0 (0.0%)	SHUNT, LP INSERTION/REVISION/REMOVAL/RM25, NEURO	1 (0.1%)	0 (0.0%)
SHUNT, VP INSERTION/REVISION/REMOVAL/MRT, NEURO	1 (0.1%)	0 (0.0%)	SIALENDOSCOPY, ORL	2 (0.2%)	0 (0.0%)
SIGMOIDOSCOPY WITH BOTOX, GI	1 (0.1%)	0 (0.0%)	SIGMOIDOSCOPY, GI	11 (0.9%)	1 (0.2%)
SKIN GRAFT, PLASTICS	8 (0.6%)	0 (0.0%)	SOFT TISSUE BIOPSY, ORL	1 (0.1%)	1 (0.2%)
SOFT TISSUE EXCISION, COMPLEX, GENSURG	20 (1.6%)	0 (0.0%)	SOFT TISSUE EXCISION, SIMPLE, GENSURG	7 (0.5%)	0 (0.0%)
SOFT TISSUE, BENIGN TUMOR EXCISION , COMPLEX, ORTHO	10 (0.8%)	0 (0.0%)	SOFT TISSUE, BENIGN TUMOR EXCISION, SIMPLE, ORTHO	3 (0.2%)	0 (0.0%)
SOFT TISSUE, MALIGNANCY RADICAL RESECTION, ORTHO	11 (0.9%)	2 (0.5%)	SPHINCTER PHARYNGOPLASTY, PLASTICS	7 (0.5%)	0 (0.0%)
SPICA CAST, APPLICATION/CHANGE, ORTHO	1 (0.1%)	0 (0.0%)	SPINAL CORD TUMOR COMPLEX/RM 25, NEURO	1 (0.1%)	0 (0.0%)
SPINAL CORD TUMOR, NEURO	17 (1.3%)	1 (0.2%)	SPINE EXPLORATION, NEURO	4 (0.3%)	0 (0.0%)
SPINE EXPLORATION, ORTHO	1 (0.1%)	0 (0.0%)	SPINE FUSION ANTERIOR, THORACOSCOPIC (VATS) WITH POSTERIOR SPINE FUSION, ORTHO	3 (0.2%)	0 (0.0%)
SPINE FUSION POSTERIOR WITH HEMIVERTEBRA EXCISION, ORTHO	4 (0.3%)	0 (0.0%)	SPINE FUSION POSTERIOR, LUMBAR TO SACRAL IN SITU, ORTHO	5 (0.4%)	0 (0.0%)
SPINE FUSION WITH NEURO DECOMPRESSION, ORTHO	4 (0.3%)	0 (0.0%)	SPINE FUSION WITH OSTEOTOMIES, ORTHO	1 (0.1%)	0 (0.0%)
SPINE FUSION WITH THORACOPLASTY, ORTHO	1 (0.1%)	0 (0.0%)	SPINE FUSION, CERVICAL, ANTERIOR, NEURO	1 (0.1%)	0 (0.0%)
SPINE PLUS: SPINE, OSTEOTOMY, ORTHO	2 (0.2%)	1 (0.2%)	SPINE PLUS: SPINE, VERTEBRAL COLUMN RESECTION, ORTHO	2 (0.2%)	0 (0.0%)
SPINE RELEASE ANTERIOR, THORACOSCOPIC (VATS), ORTHO	1 (0.1%)	0 (0.0%)	SPINE TUMOR, NEURO	1 (0.1%)	3 (0.7%)
SPINE VERTEBRAL COLUMN RESECTION, ORTHO	5 (0.4%)	0 (0.0%)	SPINE, ANTERIOR, OPEN, ORTHO	5 (0.4%)	0 (0.0%)
SPINE, ANTERIOR, THORACOSCOPIC W/VATS, ORTHO	7 (0.5%)	1 (0.2%)	SPINE, CAST BODY APPLICATION, ORTHO	5 (0.4%)	0 (0.0%)
SPINE, FUSION W/PLIF, ORTHO	13 (1.0%)	3 (0.7%)	SPINE, FUSION, ANTERIOR, CERVICAL, ORTHO	5 (0.4%)	2 (0.5%)
SPINE, FUSION, POSTERIOR, CERVICAL TO THORACIC, ORTHO	6 (0.5%)	0 (0.0%)	SPINE, FUSION, POSTERIOR, CERVICAL, ORTHO	8 (0.6%)	2 (0.5%)
SPINE, FUSION, POSTERIOR, LUMBAR TO SACRAL, ORTHO	42 (3.3%)	6 (1.4%)	SPINE, FUSION, POSTERIOR, OCCIPUT TO CERVICAL TO THORACIC, ORTHO	1 (0.1%)	0 (0.0%)
SPINE, FUSION, POSTERIOR, OCCIPUT TO CERVICAL, ORTHO	3 (0.2%)	3 (0.7%)	SPINE, FUSION, POSTERIOR, THORACIC TO LUMBAR, ORTHO	263 (20.7%)	46 (10.4%)
SPINE, FUSION, POSTERIOR, THORACIC TO PELVIS, ORTHO	83 (6.5%)	7 (1.6%)	SPINE, FUSION, POSTERIOR, THORACIC, ORTHO	346 (27.2%)	32 (7.2%)
SPINE, FUSION, REVISION, POSTERIOR, ORTHO	33 (2.6%)	9 (2.0%)	SPINE, GROWING ROD EXCHANGE, ORTHO	21 (1.6%)	8 (1.8%)
SPINE, GROWING ROD INSERTION, ORTHO	39 (3.1%)	1 (0.2%)	SPINE, GROWING ROD LENGTHENING, ORTHO	21 (1.6%)	4 (0.9%)
SPINE, HALO APPLICATION, ORTHO	9 (0.7%)	0 (0.0%)	SPINE, HARDWARE REMOVAL, ORTHO	35 (2.7%)	3 (0.7%)
SPINE, INCISION AND DRAINAGE, ORTHO	2 (0.2%)	0 (0.0%)	SPINE, STERNOCLEIDOMASTOID RELEASE, ORTHO	4 (0.3%)	0 (0.0%)
SPINE, VEPTR INSERTION, ORTHO	6 (0.5%)	0 (0.0%)	SPINE, VEPTR LENGTHENING, ORTHO	6 (0.5%)	1 (0.2%)
SPLENECTOMY, COMPLETE, LAPAROSCOPIC, 5MM, GENSURG	22 (1.7%)	2 (0.5%)	SPLENECTOMY, OPEN, GENSURG	1 (0.1%)	0 (0.0%)
SPLENECTOMY, PARTIAL, LAPAROSCOPIC, 5MM, GENSURG	2 (0.2%)	1 (0.2%)	SPLIT THICKNESS SKIN GRAFT, PLASTICS	3 (0.2%)	0 (0.0%)
STEP PROCEDURE, 9 TO 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	STEREOTACTIC BRAIN BIOPSY, NEURO	2 (0.2%)	0 (0.0%)
STEREOTACTIC BRAIN BIOPSY/MRT, NEURO	2 (0.2%)	0 (0.0%)	STEREOTACTIC IMPLANTATION OF EPILEPSY ELECTRODES, ROBOT ASSISTED, NEURO	4 (0.3%)	1 (0.2%)
STERNAL RECONSTRUCTION, ORTHO	1 (0.1%)	0 (0.0%)	STERNOTOMY, THYMECTOMY, GENSURG	5 (0.4%)	3 (0.7%)
STOMA REVISION, GU	8 (0.6%)	0 (0.0%)	STRABISMUS 1 MUSCLE, EYE	4 (0.3%)	0 (0.0%)
STRABISMUS 1 MUSCLE/REOPERATION/ADJUSTMENT IN RECOVERY WITH SEDATION, EYE	1 (0.1%)	0 (0.0%)	STRABISMUS 2 MUSCLES, EYE	11 (0.9%)	0 (0.0%)
STRABISMUS 2 MUSCLES/ADJUSTMENT IN RECOVERY WITH SEDATION, EYE	5 (0.4%)	0 (0.0%)	STRABISMUS 2 MUSCLES/ADJUSTMENT IN RECOVERY-NO SEDATION, EYE	1 (0.1%)	0 (0.0%)

Table A.2: Summary statistics of primary procedure

Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)	Primary Procedure, no. %	Training data (n=19281)	Test data (n=1777)
STRABISMUS 2 MUSCLES/REOPERATION, EYE	2 (0.2%)	0 (0.0%)	STRABISMUS 3 OR MORE MUSCLES, EYE	1 (0.1%)	0 (0.0%)
STRABISMUS 3 OR MORE MUSCLES/ADJUSTMENT IN RECOVERY WITH SEDATION, EYE	1 (0.1%)	0 (0.0%)	STRABISMUS 3 OR MORE MUSCLES/REOPERATION/ADJUSTMENT IN RECOVERY WITH SEDATION, EYE	1 (0.1%)	1 (0.2%)
STRICTUROPLASTY, GENSURG	5 (0.4%)	0 (0.0%)	STUMP REVISION, ORTHO	4 (0.3%)	0 (0.0%)
SUB-GALEAL SHUNT (SGS) PLACEMENT, NEURO	1 (0.1%)	0 (0.0%)	SUBLINGUAL GLAND EXCISION, ORL	14 (1.1%)	2 (0.5%)
SUBMANDIBULAR GLAND EXCISION, ORL	13 (1.0%)	1 (0.2%)	SUBTALAR BAR RESECTION, ORTHO	8 (0.6%)	0 (0.0%)
SUPPRELIN IMPLANT REPLACEMENT, GU	1 (0.1%)	0 (0.0%)	SUPRAPUBIC/NEPHROSTOMY TUBE INSERTION/REPLACEMENT, PERCUTANEOUS, GU	2 (0.2%)	0 (0.0%)
SURGICAL HIP DISLOCATION, ORTHO	4 (0.3%)	0 (0.0%)	SUTURE REMOVAL, ENDOSCOPY, NEURO	171 (13.4%)	1 (0.2%)
SUTURE REMOVAL, GENSURG	1 (0.1%)	0 (0.0%)	SYNDACTYLY RELEASE, COMPLEX, FOOT, PLASTICS	9 (0.7%)	0 (0.0%)
SYNDACTYLY RELEASE, COMPLEX, HAND, PLASTICS	35 (2.7%)	3 (0.7%)	SYNDACTYLY RELEASE, SIMPLE, FOOT, PLASTICS	1 (0.1%)	0 (0.0%)
SYNDACTYLY RELEASE, SIMPLE, PLASTICS	11 (0.9%)	0 (0.0%)	TALOCALCANEAL COALITION EXCISION, ORTHO	1 (0.1%)	0 (0.0%)
TARSORRHAPHY, EYE	1 (0.1%)	0 (0.0%)	TENDON LENGTHENING/RELEASE, BILATERAL, ORTHO	10 (0.8%)	2 (0.5%)
TENDON LENGTHENING/RELEASE, UNILATERAL, ORTHO	10 (0.8%)	0 (0.0%)	TENDON RELEASE/TENOTOMY/TENOLYSIS, PLASTICS	2 (0.2%)	0 (0.0%)
TENDON REPAIR ANKLE/FOOT, ORTHO	2 (0.2%)	0 (0.0%)	TENDON TRANSFER TIBIAL, ORTHO	1 (0.1%)	0 (0.0%)
TESTICULAR IMPLANT INSERTION, GU	1 (0.1%)	0 (0.0%)	TETHERED CORD RELEASE COMPLEX, NEURO	113 (8.9%)	17 (3.8%)
TETHERED CORD RELEASE COMPLEX/RM 25, NEURO	6 (0.5%)	0 (0.0%)	TETHERED CORD RELEASE/FATTY FILUM, NEURO	100 (7.9%)	15 (3.4%)
THIRD VENTRICULOSTOMY (ETV) AND CHOROID PLEXUS CAUTERIZATION (CPC), ENDOSCOPY, NEURO	23 (1.8%)	6 (1.4%)	THIRD VENTRICULOSTOMY (ETV) WITH BRAIN BIOPSY, ENDOSCOPY, NEURO	2 (0.2%)	0 (0.0%)
THIRD VENTRICULOSTOMY (ETV), ENDOSCOPY, NEURO	41 (3.2%)	5 (1.1%)	THORACIC OUTLET DECOMPRESSION, GENSURG	25 (2.0%)	10 (2.3%)
THORACOSCOPY, DIAGNOSTIC, 3MM, GENSURG	1 (0.1%)	0 (0.0%)	THORACOSCOPY, DIAGNOSTIC, 5MM, GENSURG	18 (1.4%)	3 (0.7%)
THORACOSCOPY, DIAGNOSTIC, ¿ 25 KG, GENSURG	1 (0.1%)	0 (0.0%)	THORACOTOMY, 9 TO 25 KG, GENSURG	18 (1.4%)	0 (0.0%)
THORACOTOMY, ¿ 25 KG, GENSURG	10 (0.8%)	0 (0.0%)	THORACOTOMY, LUNG LOBECTOMY, GENSURG	7 (0.5%)	0 (0.0%)
THORACOTOMY, LUNG RESECTION, 9 TO 25 KG, GENSURG	14 (1.1%)	1 (0.2%)	THORACOTOMY, LUNG RESECTION, ¡ 9 KG, GENSURG	1 (0.1%)	1 (0.2%)
THORACOTOMY, LUNG RESECTION, ¿ 25 KG, GENSURG	4 (0.3%)	9 (2.0%)	THORACOTOMY, MEDIASTINAL MASS RESECTION, 9 TO 25 KG, GENSURG	6 (0.5%)	0 (0.0%)
THORACOTOMY, PNEUMONECTOMY, GENSURG	1 (0.1%)	0 (0.0%)	THYMECTOMY, THORACOSCOPIC, 5MM, GENSURG	2 (0.2%)	0 (0.0%)
THYROGLOSSAL DUCT CYST EXCISION , ORL	48 (3.8%)	4 (0.9%)	THYROGLOSSAL DUCT CYST EXCISION, GENSURG	7 (0.5%)	0 (0.0%)
THYROIDECTOMY WITH NODE DISSECTION, GENSURG	3 (0.2%)	3 (0.7%)	THYROIDECTOMY, COMPLETION LOBECTOMY, GENSURG	6 (0.5%)	0 (0.0%)
THYROIDECTOMY, LOBECTOMY, GENSURG	46 (3.6%)	10 (2.3%)	THYROIDECTOMY, ORL	3 (0.2%)	0 (0.0%)
THYROIDECTOMY, SUBTOTAL, FOR GRAVE'S DISEASE, GENSURG	6 (0.5%)	0 (0.0%)	THYROIDECTOMY, TOTAL FOR MALIGNANT NODULE, GENSURG	3 (0.2%)	0 (0.0%)
THYROIDECTOMY, TOTAL, WITH LYMPH NODE DISSECTION, GENSURG	12 (0.9%)	0 (0.0%)	THYROIDECTOMY, TOTAL/SUBTOTAL, GENSURG	40 (3.1%)	5 (1.1%)
TIBIA, HARDWARE REMOVAL OF LOCKING IM NAIL/IM ROD, ORTHO	1 (0.1%)	0 (0.0%)	TIBIA, OSTEOTOMY WITH PUDDU PLATE, ORTHO	1 (0.1%)	0 (0.0%)
TIBIA/FIBULA JOINT, LIGAMENT RECONSTRUCTION, ORTHO	2 (0.2%)	1 (0.2%)	TIBIA/FIBULA, ORIF, ORTHO	22 (1.7%)	2 (0.5%)
TIBIA/FIBULA, OSTEOTOMY, EXTERNAL FIXATOR, ORTHO	31 (2.4%)	0 (0.0%)	TIBIA/FIBULA, OSTEOTOMY, IM NAIL/IM ROD, ORTHO	36 (2.8%)	0 (0.0%)
TIBIA/FIBULA, OSTEOTOMY, NAIL, ORTHO	3 (0.2%)	0 (0.0%)	TIBIA/FIBULA, OSTEOTOMY, PLATE/SCREW, ORTHO	21 (1.6%)	7 (1.6%)
TISSUE EXPANDER INJECTION, GENSURG	1 (0.1%)	0 (0.0%)	TISSUE EXPANDER INSERTION, PLASTICS	8 (0.6%)	0 (0.0%)
TISSUE EXPANDER REMOVAL, COMPLEX RECONSTRUCTION, PLASTICS	3 (0.2%)	1 (0.2%)	TISSUE EXPANDER REMOVAL, SIMPLE RECONSTRUCTION, PLASTICS	1 (0.1%)	0 (0.0%)
TMJ ARTHROPLASTY, UNILATERAL, PLASTICS	8 (0.6%)	0 (0.0%)	TMJ CONSTRUCTION, COSTOCHONDRAL GRAFT, UNILATERAL, PLASTICS	1 (0.1%)	0 (0.0%)
TMJ MANIPULATION, PLASTICS	3 (0.2%)	0 (0.0%)	TMJ REPLACEMENT OR CONSTRUCTION, COMPLEX ¿5 HOURS, PLASTICS	1 (0.1%)	0 (0.0%)
TMJ REPLACEMENT, ALLOPLASTIC, BILATERAL, PLASTICS	13 (1.0%)	1 (0.2%)	TMJ REPLACEMENT, ALLOPLASTIC, UNILATERAL, PLASTICS	1 (0.1%)	2 (0.5%)
TMJ REPLACEMENT, ALLOPLASTIC, WITH ORTHOGNATHIC SURGERY, PLASTICS	1 (0.1%)	1 (0.2%)	TOE TRANSFER, PLASTICS	1 (0.1%)	0 (0.0%)
TONGUE BASE EXCISION, ORL	5 (0.4%)	0 (0.0%)	TONSILLECTOMY WITH ADENOIDECTOMY, ORL	1809 (142.1%)	122 (27.6%)
TONSILLECTOMY, LINGUAL, ORL	19 (1.5%)	2 (0.5%)	TONSILLECTOMY, ORL	103 (8.1%)	11 (2.5%)
TONSILLOTOMY WITH ADENOIDECTOMY, ORL	599 (47.1%)	48 (10.9%)	TONSILLOTOMY, ORL	32 (2.5%)	3 (0.7%)
TOTAL CALVARIAL VAULT REMODEL, PLASTICS	25 (2.0%)	2 (0.5%)	TRABECULECTOMY WITH MITOMYCIN, EYE	2 (0.2%)	0 (0.0%)
TRABECULOTOMY, EYE	2 (0.2%)	0 (0.0%)	TRACHEAL STOMA REVISION/CLOSURE, ORL	2 (0.2%)	0 (0.0%)
TRACHEOCUTANEOUS FISTULA REPAIR, ORL	13 (1.0%)	2 (0.5%)	TRACHEOESOPHAGEAL FISTULA REPAIR, H TYPE, GENSURG	3 (0.2%)	0 (0.0%)
TRACHEOESOPHAGEAL FISTULA REPAIR, OPEN, GENSURG	6 (0.5%)	0 (0.0%)	TRACHEOPEXY, ROBOTIC, GENSURG	5 (0.4%)	0 (0.0%)
TRACHEOSTOMY, ORL	7 (0.5%)	0 (0.0%)	TRANSBRONCHIAL BIOPSY, PULMONARY	2 (0.2%)	0 (0.0%)
TRANSSPHENOIDAL BIOPSY/HYPOPHYSECTOMY, NEURO	24 (1.9%)	6 (1.4%)	TRANSSPHENOIDAL RESECTION OF PITUITARY TUMOR, NEURO	4 (0.3%)	0 (0.0%)
TRANSSPHENOIDAL RESECTION OF PITUITARY TUMOR/RM 25, NEURO	1 (0.1%)	0 (0.0%)	TRANSURETHRAL RESECTION OF BLADDER TUMOR, GU	1 (0.1%)	0 (0.0%)
TRIGGER THUMB RELEASE, PLASTICS	1 (0.1%)	0 (0.0%)	TUMOR BIOPSY, ENDOSCOPY, NEURO	2 (0.2%)	0 (0.0%)
TURBINATE REDUCTION, COBLATOR, ORL	20 (1.6%)	2 (0.5%)	TURBINATE REDUCTION, MICRODEBRIDER, ORL	4 (0.3%)	1 (0.2%)
TYMPANOMASTOIDECTOMY, BASIC, ORL	10 (0.8%)	1 (0.2%)	TYMPANOMASTOIDECTOMY, COMPLEX, ORL	48 (3.8%)	2 (0.5%)
TYMPANOPLASTY, POSTAURICULAR, BASIC, ORL	21 (1.6%)	2 (0.5%)	TYMPANOPLASTY, POSTAURICULAR, COMPLEX, ORL	53 (4.2%)	3 (0.7%)
TYMPANOPLASTY, TRANSCANAL, BASIC, ORL	2 (0.2%)	0 (0.0%)	TYMPANOPLASTY, TRANSCANAL, COMPLEX, ORL	16 (1.3%)	1 (0.2%)
TYMPANOPLASTY, TRANSCANAL, ENDOSCOPIC, BASIC, ORL	2 (0.2%)	0 (0.0%)	TYMPANOPLASTY, TRANSCANAL, ENDOSCOPIC, COMPLEX, ORL	4 (0.3%)	0 (0.0%)
TYMPANOSTOMY TUBE INSERTION, ORL	159 (12.5%)	12 (2.7%)	TYMPANOSTOMY, ORL	2 (0.2%)	0 (0.0%)
ULCER EXCISION, ISCHIECTOMY, PLASTICS	2 (0.2%)	0 (0.0%)	UMBILICAL GRANULOMA EXCISION, GENSURG	1 (0.1%)	0 (0.0%)
UPPER EXTREMITY, HARDWARE REMOVAL, DISTAL, ORTHO	6 (0.5%)	0 (0.0%)	UPPER EXTREMITY, INFIX, ORTHO	1 (0.1%)	0 (0.0%)
UPPER EXTREMITY, PERIPHERAL NERVE NEUROPLASTY/REPAIR/NERVE GRAFTING, ORTHO	8 (0.6%)	0 (0.0%)	URACHAL CYST/SINUS EXCISION, GENSURG	3 (0.2%)	0 (0.0%)
URETERAL REIMPLANT WITH MEGAURETER REPAIR, GU	23 (1.8%)	7 (1.6%)	URETERAL REIMPLANT WITH MEGAURETER TAPERING, GU	17 (1.3%)	0 (0.0%)
URETERAL REIMPLANT, BILATERAL, GU	124 (9.7%)	14 (3.2%)	URETERAL REIMPLANT, BILATERAL/CYSTOSCOPY, GU	135 (10.6%)	16 (3.6%)
URETERAL REIMPLANT, UNILATERAL, GU	31 (2.4%)	5 (1.1%)	URETERAL REIMPLANT, UNILATERAL/CYSTOSCOPY, GU	152 (11.9%)	19 (4.3%)
URETERAL, CATHETERIZABLE CHANNEL, GU	2 (0.2%)	0 (0.0%)	URETEROSTOMY CREATION/CLOSURE, GU	2 (0.2%)	0 (0.0%)
URETEROURETEROSTOMY/TRANSURETEROURETEROSTOMY, GU	1 (0.1%)	0 (0.0%)	URETHRAL FISTULA REPAIR, PENILE, GU	10 (0.8%)	0 (0.0%)
URETHRAL FISTULA REPAIR, VAGINAL, GU	1 (0.1%)	0 (0.0%)	URETHRAL LENGTHENING, GENDER REASSIGNMENT, GU	3 (0.2%)	0 (0.0%)
URETHROPLASTY, ANTERIOR, GU	11 (0.9%)	5 (1.1%)	URETHROPLASTY, COMPLEX, GU	3 (0.2%)	0 (0.0%)
URETHROPLASTY, PERINEAL, GU	11 (0.9%)	2 (0.5%)	URINARY CATHETER INSERTION, MEDICAL	17 (1.3%)	0 (0.0%)
UROGENITAL SINUS ANOMALY REPAIR, GU	2 (0.2%)	0 (0.0%)	UVULOPALATOPHARYNGOPLASTY, ORL	2 (0.2%)	1 (0.2%)
VAC DRESSING APPLICATION/CHANGE, PLASTICS	3 (0.2%)	0 (0.0%)	VAGAL NERVE STIMULATOR INSERTION/REVISION/REMOVAL, NEURO	7 (0.5%)	0 (0.0%)
VAGINAL HYSTERECTOMY WITH LAPAROSCOPIC ASSISTANCE, GYN	23 (1.8%)	6 (1.4%)	VAGINECTOMY, GU	1 (0.1%)	0 (0.0%)
VAGINOPLASTY FOR AGENESIS, LOWER VAGINA (PULLTHROUGH), GYN	3 (0.2%)	0 (0.0%)	VAGINOPLASTY, COMPLEX, GU	7 (0.5%)	0 (0.0%)
VAGINOPLASTY, GENDER REASSIGNMENT, GU	11 (0.9%)	0 (0.0%)	VAGINOPLASTY, GENSURG	2 (0.2%)	2 (0.5%)
VAGINOPLASTY, GU	9 (0.7%)	1 (0.2%)	VAGINOPLASTY, LONG, GYN	3 (0.2%)	1 (0.2%)
VAGINOPLASTY, ROBOTIC, GENSURG	2 (0.2%)	0 (0.0%)	VAGINOPLASTY, SHORT, GYN	1 (0.1%)	0 (0.0%)
VAGINOSCOPY, 9 TO 25 KG, GENSURG	2 (0.2%)	0 (0.0%)	VAGINOSCOPY, GU	2 (0.2%)	0 (0.0%)
VAGINOSCOPY, GYN	1 (0.1%)	0 (0.0%)	VASCULAR FREE TISSUE TRANSFER, FREE FIBULA, ORTHO	1 (0.1%)	0 (0.0%)
VASCULAR MALFORMATION EXCISION, GREATER THAN 5 CM, ORTHO	53 (4.2%)	3 (0.7%)	VASCULAR MALFORMATION EXCISION, LESS THAN 5 CM, ORTHO	8 (0.6%)	0 (0.0%)
VASCULAR RING REPAIR, GENSURG	1 (0.1%)	0 (0.0%)	VENTRICULAR PERITONEAL (VP) SHUNT, NEURO	56 (4.4%)	6 (1.4%)
VEPTR INSERTION, ORTHO	1 (0.1%)	0 (0.0%)	VEPTR REMOVAL, ORTHO	4 (0.3%)	0 (0.0%)
VEPTR REVISION, ORTHO	1 (0.1%)	0 (0.0%)	VESICOSTOMY CREATION/CLOSURE, GU	9 (0.7%)	0 (0.0%)
VESTIBULOPLASTY, PLASTICS	1 (0.1%)	0 (0.0%)	VITRECTOMY, COMPLEX, EYE	3 (0.2%)	0 (0.0%)
VITRECTOMY/CAPSULOTOMY, ANTERIOR, EYE	1 (0.1%)	0 (0.0%)	WOUND CLOSURE, COMPLEX, PLASTICS	1 (0.1%)	0 (0.0%)
WOUND CLOSURE, SIMPLE, PLASTICS	1 (0.1%)	0 (0.0%)	WOUND REVISION, NEURO	5 (0.4%)	0 (0.0%)
WRIST, ARTHROSCOPY, ORTHO	3 (0.2%)	0 (0.0%)	WRIST, ARTHROSCOPY, TFCC REPAIR (NOT OPEN), ORTHO	4 (0.3%)	0 (0.0%)
WRIST, ARTHROSCOPY, TFCC REPAIR (OPEN), ORTHO	2 (0.2%)	0 (0.0%)	WRIST, EPIPHYSIODESIS, ORTHO	1 (0.1%)	0 (0.0%)
WRIST, FUSION, ORTHO	5 (0.4%)	1 (0.2%)	WRIST, GANGLION EXCISION, ORTHO	2 (0.2%)	0 (0.0%)
WRIST, OSTEOTOMY, ORTHO	10 (0.8%)	1 (0.2%)	WRIST, RECONSTRUCTION, ORTHO	2 (0.2%)	0 (0.0%)
zzABLATION THERMAL LASER, MRI CONTROLLED/MRT, NEURO	4 (0.3%)	0 (0.0%)	zzABLATION THERMAL LASER, MRI CONTROLLED/RM 25, NEURO	1 (0.1%)	0 (0.0%)
zzBACLOFEN PUMP INSERTION/RM 25, NEURO	9 (0.7%)	0 (0.0%)	zzCRANIOTOMY, ANEURYSM/AVM SIMPLE/RM 25, NEURO	2 (0.2%)	0 (0.0%)
zzCRANIOTOMY, BIOPSY/MRT, NEURO	3 (0.2%)	0 (0.0%)	zzCRANIOTOMY, FRONTAL, TUMOR/MRT, NEURO	3 (0.2%)	0 (0.0%)
zzCRANIOTOMY, FRONTAL, TUMOR/RM 25, NEURO	1 (0.1%)	0 (0.0%)	zzCRANIOTOMY, MOYA-MOYA BILATERAL/RM 25, NEURO	7 (0.5%)	0 (0.0%)
zzCRANIOTOMY, OCCIPITAL, TUMOR, NEURO	1 (0.1%)	0 (0.0%)	zzCRANIOTOMY, OCCIPITAL, TUMOR/RM 25, NEURO	2 (0.2%)	0 (0.0%)
zzCRANIOTOMY, POSTERIOR FOSSA, NEURO	2 (0.2%)	0 (0.0%)	zzCRANIOTOMY, RESECTION OF SEIZURE FOCUS/RM 25, NEURO	1 (0.1%)	0 (0.0%)
zzCRANIOTOMY, RESECTION OF TEMPORAL LOBE, NEURO	3 (0.2%)	0 (0.0%)	zzCRANIOTOMY, RESECTION OF TEMPORAL LOBE/RM 25, NEURO	1 (0.1%)	0 (0.0%)
zzCRANIOTOMY, SUBOCCIPITAL, TUMOR SIMPLE/MRT, NEURO	1 (0.1%)	0 (0.0%)	zzCRANIOTOMY, SUBOCCIPITAL, TUMOR SIMPLE/RM 25, NEURO	1 (0.1%)	0 (0.0%)
zzCRANIOTOMY, TEMPORAL, TUMOR/MRT, NEURO	1 (0.1%)	0 (0.0%)	zzCRANIOTOMY, TEMPORAL, TUMOR/RM 25, NEURO	2 (0.2%)	0 (0.0%)
zzCRANIOTOMY, TUMOR COMPLEX/RM 25, NEURO	2 (0.2%)	0 (0.0%)	zzCRANIOTOMY, TUMOR SIMPLE/MRT, NEURO	6 (0.5%)	0 (0.0%)
zzCRANIOTOMY/MRT, NEURO	1 (0.1%)	0 (0.0%)	zzCRANIOTOMY/RM 25, NEURO	2 (0.2%)	0 (0.0%)
zzDERMOID CYST EXCISION/RM 25, NEURO	1 (0.1%)	0 (0.0%)	zzENDOSCOPY FOR BRAIN BIOPSY/RM 25, NEURO	1 (0.1%)	0 (0.0%)
zzENDOSCOPY FOR FENESTRATION OF CYST, RM 25, NEURO	1 (0.1%)	0 (0.0%)	zzINTRACRANIAL CATHETER FOR CHEMOTHERAPY/MRT, NEURO	1 (0.1%)	0 (0.0%)
zzLAMINECTOMY, SIMPLE, NEURO	4 (0.3%)	0 (0.0%)	zzLAMINECTOMY/RM 25, COMPLEX, NEURO	3 (0.2%)	0 (0.0%)
zzLAMINECTOMY/RM 25, SIMPLE, NEURO	1 (0.1%)	0 (0.0%)	zzLIPOMYLOMENINGOCELE REPAIR, NEURO	1 (0.1%)	0 (0.0%)
zzPERCUTANEOUS NEPHROLITHOTOMY WITH CYBERWAND, GU	1 (0.1%)	0 (0.0%)	zzPOSTERIOR TRACHEOPLEXY, GEN SURG	1 (0.1%)	0 (0.0%)
zzROBOT ASSISTED ABLATION THERMAL LASER, MRI CONTROLLED, NEURO	1 (0.1%)	0 (0.0%)	zzSHUNT, VP INSERTION/REVISION/REMOVAL/RM 25, NEURO	5 (0.4%)	0 (0.0%)
zzSPINAL CORD TUMOR SIMPLE, NEURO	2 (0.2%)	0 (0.0%)	zzSPINE FUSION POSTERIOR, LUMBAR, ORTHO	13 (1.0%)	0 (0.0%)
zzSTEREOTACTIC BRAIN BIOPSY/RM 25, NEURO	3 (0.2%)	0 (0.0%)	zzTETHERED CORD RELEASE SIMPLE, NEURO	16 (1.3%)	0 (0.0%)
zzTETHERED CORD RELEASE SIMPLE/RM 25, NEURO	1 (0.1%)	0 (0.0%)	zzTETHERED CORD RELEASE/FATTY FILUM/RM 25, NEURO	2 (0.2%)	0 (0.0%)
zzVAGAL NERVE STIMULATOR INSERTION/REVISION/REMOVAL/RM 25, NEURO	1 (0.1%)	0 (0.0%)			

Bibliography

[1] R. Schmidt, S. Geisler, and C. Spreckelsen, "Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources," *BMC Med. Inform. Decis. Mak.*, vol. 13, p. 3, Jan. 2013.

[2] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, "Real-time prediction of inpatient length of stay for discharge prioritization," *J. Am. Med. Inform. Assoc.*, vol. 23, pp. e2–e10, Apr. 2016.

[3] N. Dehouche, S. Viravan, U. Santawat, N. Torsuwan, S. Taijan, A. Intharakosum, and Y. Sirivatanauksorn, "Hospital length of stay: A cross-specialty analysis and beta-geometric model," *medRxiv*, 2021.

[4] B. A. Marfil-Garza, P. F. Belaunzarán-Zamudio, A. Gulias-Herrero, A. C. Zuñiga, Y. Caro-Vega, D. Kershenobich-Stalnikowitz, and J. Sifuentes-Osornio, "Risk factors associated with prolonged hospital length-of-stay: 18-year retrospective study of hospitalizations in a tertiary healthcare center in mexico," *PLoS One*, vol. 13, p. e0207203, Nov. 2018.

[5] P.-F. J. Tsai, P.-C. Chen, Y.-Y. Chen, H.-Y. Song, H.-M. Lin, F.-M. Lin, and Q.-P. Huang, "Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural network," *Journal of Healthcare Engineering*, vol. 2016, p. 7035463, Apr 2016.

[6] C. Gao, A. N. Kho, C. Ivory, S. Osmundson, B. A. Malin, and Y. Chen, "Predicting length of stay for obstetric patients via electronic medical records," *Stud. Health Technol. Inform.*, vol. 245, pp. 1019–1023, 2017.

[7] L. Turgeman, J. H. May, and R. Sciulli, "Insights from a machine learning model for predicting the hospital length of stay (los) at the time of admission," *Expert Systems with Applications*, vol. 78, pp. 376–385, 2017.

[8] F. Bert, O. Kakaa, A. Corradi, A. Mascaro, S. Roggero, D. Corsi, A. Scarmozzino, and R. Siliquini, "Predicting length of stay and discharge destination for surgical patients: A cohort study," *Int. J. Environ. Res. Public Health*, vol. 17, p. 9490, Dec. 2020.

[9] M.-T. Chuang, Y.-H. Hu, and C. L. Lo, "Predicting the prolonged length of stay of general surgery patients: A supervised learning approach," *International Transactions in Operational Research*, vol. 25, 05 2016.

[10] V. Lequertier, T. Wang, J. Fondrevelle, V. Augusto, and A. Duclos, "Hospital length of stay prediction methods: A systematic review," *Med. Care*, vol. 59, pp. 929–938, Oct. 2021.

[11] S. Purushotham, C. Meng, Z. Che, and Y. Liu, "Benchmark of deep learning models on large healthcare MIMIC datasets," *CoRR*, vol. abs/1710.08531, 2017.

[12] K. Stone, R. Zwiggelaar, P. Jones, and N. Mac Parthaláin, *Predicting Hospital Length of Stay for Accident and Emergency Admissions*, pp. 283–295. 01 2020.

[13] D. Castiñeira, K. R. Schlosser, A. Geva, A. R. Rahmani, G. Fiore, B. K. Walsh, C. D. Smallwood, J. H. Arnold, and M. Santillana, "Adding continuous vital sign information to static clinical data improves the prediction of length of stay after intubation: A data-driven machine learning approach," *Respiratory Care*, vol. 65, no. 9, pp. 1367–1377, 2020.

[14] B. Nestor, M. B. A. McDermott, W. Boag, G. Berner, T. Naumann, M. C. Hughes, A. Goldenberg, and M. Ghassemi, "Feature robustness in non-stationary health records: Caveats to deployable model performance in common clinical machine learning tasks," *CoRR*, vol. abs/1908.00690, 2019.

[15] M. A. Abd-Elrazek, A. A. Eltahawi, M. H. Abd Elaziz, and M. N. Abd-Elwhab, "Predicting length of stay in hospitals intensive care unit using general admission features," *Ain Shams Engineering Journal*, vol. 12, no. 4, pp. 3691–3702, 2021.

[16] T. A. Daghistani, R. Elshawi, S. Sakr, A. M. Ahmed, A. Al-Thwayee, and M. H. Al-Mallah, "Predictors of in-hospital length of stay among cardiac patients: A machine learning approach," *International Journal of Cardiology*, vol. 288, pp. 140–147, 2019.

[17] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, "Use of data mining techniques to determine and predict length of stay of cardiac patients," *Healthc. Inform. Res.*, vol. 19, pp. 121–129, June 2013.

[18] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G. E. Duggan, J. Irvine, Q. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. D. Howell, C. Cui, G. S. Corrado, and J. Dean, "Scalable and accurate deep learning with electronic health records," *npj Digital Medicine*, vol. 1, p. 18, May 2018.

[19] "Current Procedural Terminology." `https://www.ama-assn.org/amaone/cpt-current-procedural-terminology`. Accessed: 2022-04-25.

[20] "Child Developmental Milestones." `https://www.cdc.gov/ncbddd/childdevelopment/positiveparenting/infants.html`. Accessed: 2022-04-25.

[21] “CPT Overview and Code Approval.” https://www.ama-assn.org/practice-management/cpt/cpt-overview-and-code-approval. Accessed: 2022-04-25.

[22] “Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses.” https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp. Accessed: 2022-04-25.

[23] K. W. Fung, “The CORE Problem List Subset of SNOMED CT.” https://confluence.ihtsdotools.org/download/attachments/73368244/SnomedCtShowcase2013_Abstract_09.pdf?version=1&modificationDate=1535031927000&api=v2. Accessed: 2022-04-26.

[24] “Inpatient vs. Observation Status Frequently Asked Questions.” https://www.upmc.com/-/media/upmc/patients-visitors/patient-info/documents/status-faq.pdf. Accessed: 2022-04-26.

[25] “Inpatient versus Observation Care.” https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/about-ama/councils/Council%20Reports/council-on-medical-service/issue-brief-inpatient-v-observation-care.pdf. Accessed: 2022-04-26.

[26] “1.4. Support Vector Machines – scikit-learn 1.0.2 documentation.” https://scikit-learn.org/stable/modules/svm.html. Accessed: 2022-04-26.

[27] J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, “Tackling the poor assumptions of naive bayes text classifiers,” in *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML’03, p. 616–623, AAAI Press, 2003.

[28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective.* The MIT Press, 2012.

[29] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[30] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, p. 5–32, oct 2001.

[31] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, (New York, NY, USA), p. 161–168, Association for Computing Machinery, 2006.

[32] "1.11 Ensemble Methods – scikit-learn 1.0.2 documentation." `https://scikit-learn.org/stable/modules/ensemble.html#voting-classifier`. Accessed: 2022-04-26.

[33] M. J. van der Laan, E. C. Polley, and A. E. Hubbard, "Super learner," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, p. Article25, Sept. 2007.

[34] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, p. 647–665, dec 2014.